The Effects of Selective Feedback on Visual Search Target Detection

Ashten de Haan

A report submitted as a partial requirement for the degree of Bachelor of Psychological Science with Honours in Psychology at the University of Tasmania, 2019

Statement of Sources

I declare that this report is my own original work and that contributions of others have been duly acknowledged

Signed: _____ Date: 17/10/2019

Acknowledgments

I would like to say a huge thank you to my supervisor Dr Matthew Palmer, for the continuous enthusiasm, guidance and support you have provided throughout the year. I really appreciate all the time and effort you have put into assisting me with all aspects of completing this thesis. I also thank you for your endless patience with my long lists of questions and my unplanned office drop ins. You do not know how much all of this helped my stress levels and sanity!

Thank you to Talira Kucina for preparing the software for this project, as well as your willingness to help out wherever you could. I have really appreciated it.

I would also like to thank all the people who gave up their time to participate in this study, especially since many of them made the trip to the university just to help me out. Without all of these people, this thesis would not be what it is.

A massive thank you to my parents for the endless support and encouragement you have provided me, not just this year, but through all my years at university. I am grateful for the numerous hours you have spent reading my work or quizzing me, as well as always being there to help me through my university related freak outs. Thanks for getting me through the year.

A final thank you goes to Reo, for your patience, understanding and constant positivity that has kept me sane these past eight or so months. Thanks for always being the voice or reason.

Table of Contents

List of Tables

List of Figures

The Effects of Selective Feedback on Visual Search Target Detection

Ashten de Haan

Word Count: 9988

Abstract

The current study investigated the effect of a retraining intervention that utilises selective feedback, in which feedback is only provided for misses, on target detection performance in low prevalence settings. Fifty-one participants (34 female) aged 18-50 years (M =26.57 years, SD =6.19 years) were randomly allocated to either the control condition, where they received no feedback, the full feedback condition, where they received feedback on all training trials, or the selective feedback condition, where they only received feedback when they missed a target on a training trial. Participants completed a visual search task that required them to look for a knife in x-ray images of luggage. The study consisted of alternating long, low prevalence 'on the job' blocks and brief, higher prevalence 'training' blocks. As expected, the selective feedback retraining procedure led to a significantly less conservative criterion than full feedback and no feedback. This translated into a significantly greater proportion of false alarms than full feedback and no feedback, and a meaningful, but non-significant, increase in hits compared to the control condition. Unexpectedly, however, there was no effect of feedback on response times. Such findings provide valuable information for the low prevalence setting of airport luggage screening.

Visual search tasks are involved in many components of daily life (Biggs, Adamo, & Mitroff, 2014). Such tasks require observers to identify targets positioned amongst numerous distractors (Wolfe, Brunelli, Rubinstein, & Horrowitz, 2013). While many visual searches contain negligible consequences for failing to detect the target (Biggs et al., 2014), accurate performance on visual search tasks plays an important role in society, as visual search tasks are utilised in critical areas such as airport luggage screening and radiology (Schwark, Sandry, Macdonald, & Dolgov, 2012).

The domains of airport luggage screening and radiology are characterised by low target prevalence (Biggs et al., 2014). Low target prevalence refers to the targets of interest (i.e. weapons and explosives in airport luggage screening, and abnormal growths in radiology) being rarely present (Biggs et al., 2014; Evans, Birdwell, & Wolfe, 2013; Harris, 2002). For instance, breast cancers are only present in approximately 5 cases per 1000 mammograms (Lehman et al., 2017). Similarly, while the exact rate at which targets exist in airport luggage is unknown, it is understood that they appear with extremely low frequency (Biggs et al., 2014; Wolfe et al., 2013). The target prevalence in these settings are, therefore, markedly different from the target prevalence of visual search tasks that are employed in laboratory studies, as laboratory studies typically contain targets on 50% of trials (Wolfe, Horowitz, & Kenner, 2005). The low prevalence rate of targets in airport luggage screening and radiology raises concerns due to what has been termed the Low Prevalence Effect (LPE).

Numerous studies have investigated potential interventions to counteract the LPE, such as implementing retraining blocks that utilise feedback (Wolfe et al., 2007). The current study draws on research from the recognition memory literature

to investigate an alternative retraining intervention that utilises selective feedback (i.e. feedback that is only provided for specific responses).

The Low Prevalence Effect

The LPE was discovered by Wolfe et al. (2005) who found that low prevalence search tasks were associated with a much greater rate of targets going undetected. In their study, Wolfe et al. (2005) found that when targets were present in 50% of trials, only 7% of those trials resulted in misses (i.e. failing to detect a target when it was present). However, when target prevalence decreased to 10%, the error rate increased to 16% of trials, and when target prevalence was at 1%, errors further increased to 30% of trials, with the large majority of these errors being misses. In addition to increased misses, Wolfe et al. (2005) also identified that prevalence rate influences response time (RT). At 50% prevalence, RTs were slower for target-absent trials than target-present trials, indicating that observers were taking time to search for the target in target-absent trials before deciding whether the target was present or absent. However, at 1% prevalence, target-absent responses became faster than target-present responses, indicating that observers were no longer taking time to search for the target before deciding that it was absent. This speeded response at 1% prevalence has been suggested to contribute to the increased error rate in low prevalence search tasks as it results in observers discontinuing their search in less than the average time required to locate a target when one is present (Wolfe et al., 2005).

The LPE has proved to be quite robust and has been replicated many times (e.g. Fleck & Mitroff, 2007; Lau & Huang, 2010) with both simple (Rich et al., 2008) and complex stimuli (Wolfe et al., 2007). This effect has also been found amongst newly trained Transportation Security Officers (TSOs; individuals who

examine luggage x-rays at airport checkpoints; Wolfe et al., 2013), as well as amongst mammographers (individuals who search for breast cancers; Evans et al., 2013) and cytologists (individuals who search for cervical cancers; Evans, Tambouret, Evered, Wilbur, & Wolfe, 2011) in true clinical settings, demonstrating that professionals working in low prevalence domains are also susceptible to the LPE. This is concerning as failure to detect targets in medical and airport security settings could have serious, life threatening, consequences (Mitroff & Biggs, 2014).

Underlying Cause of the Low Prevalence Effect

There have been multiple suggestions for what causes the LPE. Laboratory studies have demonstrated that the increased misses at low prevalence is not the result of careless error. If this were the case, the errors of two observers viewing the same sequence of trials should be uncorrelated (Wolfe et al., 2007). However, Wolfe et al. (2007) found a strong correlation in the errors made between two observers viewing the same stimuli, therefore, ruling this explanation out.

Speed-Accuracy Trade-Off and Motor Errors

It has been proposed that a speed-accuracy trade-off underlies the LPE, suggesting that the increased misses at low prevalence is the result of speeded responses (Fleck & Mitroff, 2007). In their study, Fleck and Mitroff (2007) found that RTs for trials that resulted in misses were typically faster than RTs for trials that resulted in hits, suggesting a relationship between speed and accuracy, where the faster the response is, the less accurate the observer is. Their results also suggested that the increased error for faster responses is due to motor errors, as when observers where provided with the opportunity to correct their initial response, misses were often corrected resulting in no effect of prevalence on performance. This led Fleck and Mitroff (2007) to conclude that the increased misses at low prevalence occurs

due to observers forming a habit of quickly responding that no target is present, so that even when they do identify a target, the habit causes them to select the wrong key.

However, other studies have challenged this explanation of the LPE. One point of interest is that Fleck and Mitroff's (2007) study used a simple search array, where items appeared from a canonical viewpoint and did not overlap. However, when correctable searches were implemented with more complex stimuli, such as busy x-ray images of luggage where items could appear from non-canonical viewpoints and overlapped, the low prevalence effect was not eliminated (Van Wert, Horowitz, & Wolfe, 2009). This, therefore, suggests that in more complex, real-world visual searches, the increased misses at low prevalence is unlikely to be due to motor errors of speeded responses.

A further finding that contradicts the notion of a speed-accuracy trade-off is Wolfe and Van Wert's (2010) finding that while RTs for target-absent responses are linked with target prevalence, target-present RTs remained relatively stable regardless of target prevalence. Wolfe and Van Wert (2010) reported that while at low prevalence target-absent RTs became faster, at high prevalence target-absent responses were slowed, indicating a positive relationship between target prevalence and target-absent RTs. They also identified that although the number of false alarms increased at high target prevalence, indicating that observers were frequently responding that the target was present, the speed of target-present responses remained relatively stable. These findings, therefore, challenge the notion of a speed-accuracy trade-off, as if this were the case, responses should be speeded at both high and low target prevalence, where it is easier to predict the outcome, and slowest at 50% target prevalence, where it is more difficult to predict the outcome (Wolfe &

Van Wert, 2010). However, this was not found as target-absent responses were speeded at low target prevalence, but target-present responses were not speeded at high target prevalence. A further point is that inducing slower responses among observers has not been found to improve accuracy on low prevalence search tasks (Wolfe et al., 2007), which is inconsistent with what would be expected if a speed-accuracy trade-off was at play (Förster, Higgins, & Bianco, 2003).

Signal Detection Theory (SDT): Criterion

Instead, the impact of target prevalence on detection is typically considered using SDT measures of *sensitivity* and *criterion* (Wolfe & Van Wert, 2010). In a standard SDT model there are assumed to be two probability distributions: one for distractors and one for targets (see Figure 1; Macmillan & Creelman, 2005). These distributions exist along a horizontal axis which represents the strength of evidence. The greater the overlap between these distributions, the greater the uncertainty surrounding whether the scene contains a target or not (Green & Swets, 1966). This relates to sensitivity, which reflects how difficult it is to differentiate a target from distractors (Macmillan & Creelman, 2005). The greater the distance between the distributions, the greater sensitivity is, and therefore, the easier it is to differentiate a target from distractors (see Figure 1b; Macmillan & Creelman, 2005). Studies investigating the LPE have determined that the increased misses in low prevalence search tasks is not due to a reduction in sensitivity, as sensitivity remains relatively stable, across different target prevalence rates (Wolfe & Van Wert, 2010).

The criterion component of SDT, on the other hand, refers to a decision-making threshold. This threshold relates to how much evidence is required to make a specific decision (see Figure 2). When evidence exceeds this threshold, a ‘yes’ response is given, indicating that a target has been identified. When evidence

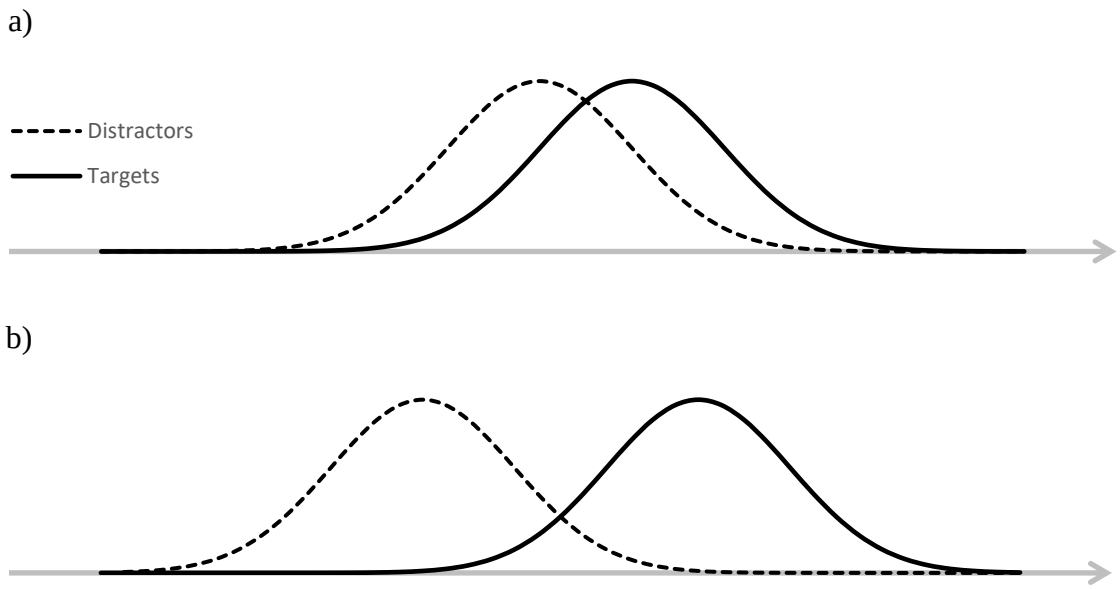


Figure 1. SDT distributions. (a) Greater overlap between the two distributions indicates low sensitivity, resulting in poorer ability to discriminate between targets and distractors. (b) Reduced overlap between the two distributions indicates high sensitivity, resulting in greater ability to discriminate between targets and distractors.

does not exceed this threshold, a ‘no’ response is given, indicating that a target has not been identified (Macmillan & Creelman, 2005). This criterion can shift, making one response more likely than the other (Macmillan & Creelman, 2005). As the criterion becomes more conservative, the observer requires more evidence to say ‘yes, a target is present’, therefore, biasing them towards saying ‘no, a target is not present’. Theoretically, this increases misses and correct rejections, while reducing false alarms and hits (see Figure 3; Green & Swets, 1966). Conversely, setting a more liberal criterion results in the observer requiring less evidence for them to say ‘yes, a target is present’ therefore, reducing the number of misses and correct rejections, but increasing the number of false alarms and hits (see Figure 4; Green & Swets, 1966).

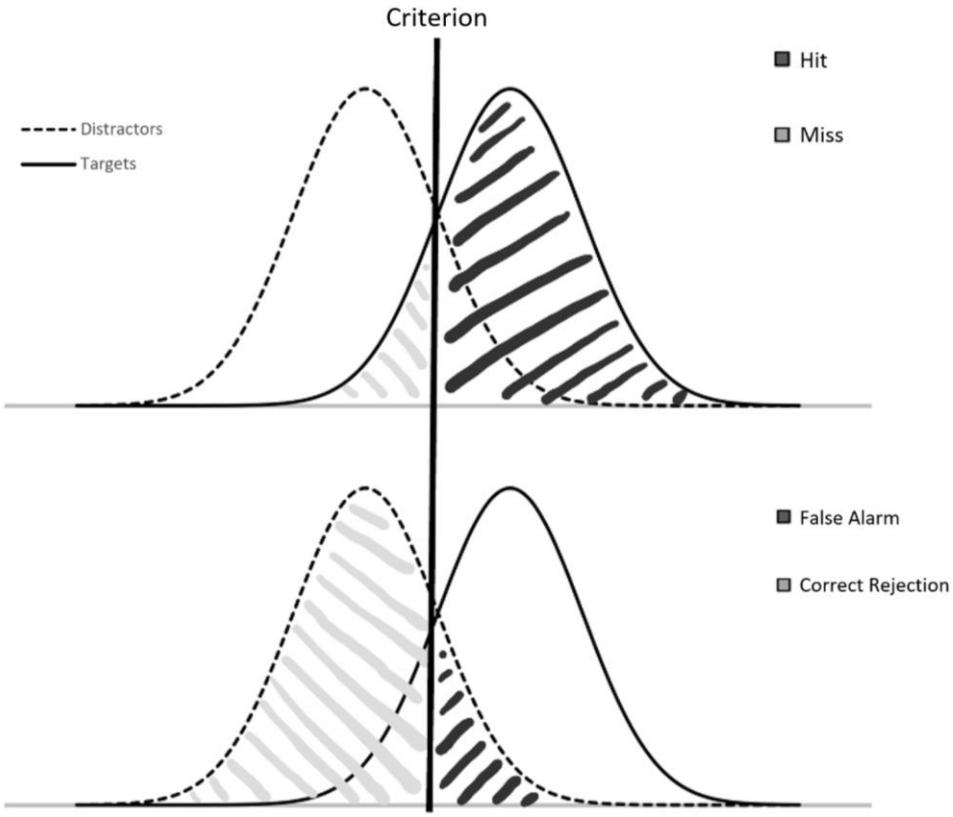


Figure 2. Neutral criterion placement.

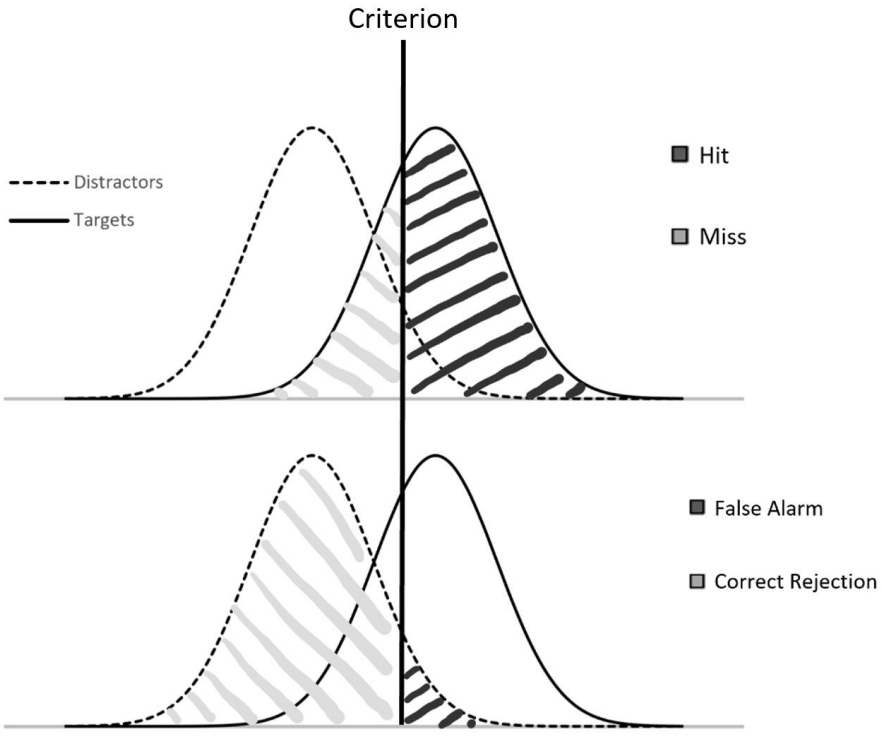


Figure 3. Conservative criterion placement.

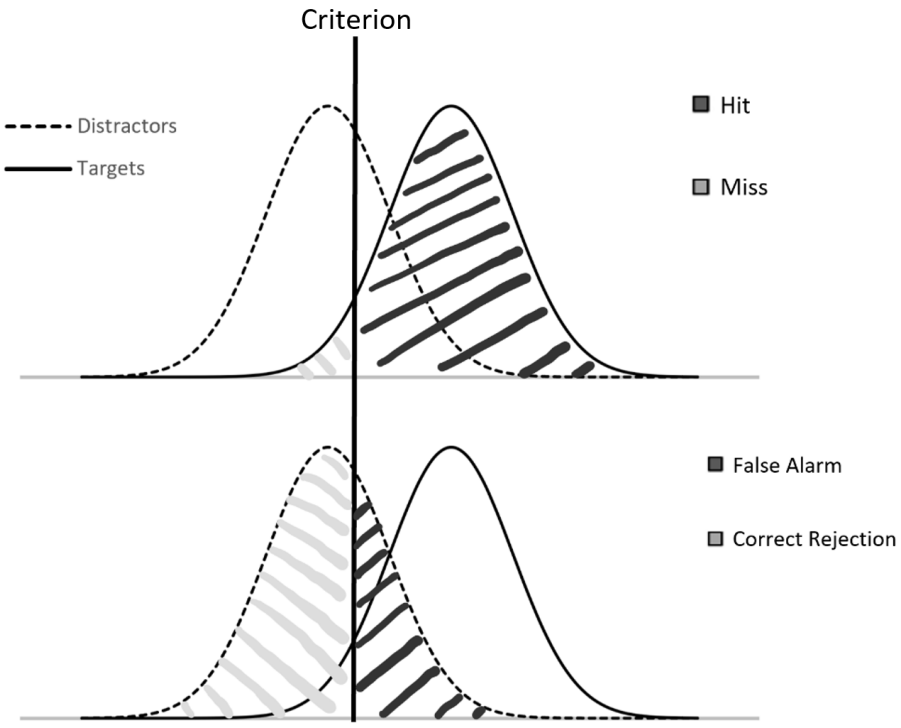


Figure 4. Liberal criterion placement.

Studies support the notion that the increased misses at low prevalence is due to a conservative shift in criterion (Wolfe et al., 2007). Wolfe et al. (2007) found observers' criteria to be reasonably neutral at 50% target prevalence. However, observer's criteria were significantly more conservative at low prevalence, therefore, biasing them towards saying 'no' (indicating that they did not detect the target), and resulting in a greater number of misses than when a neutral criterion was set. Wolfe and Van Wert (2010) also found that criterion placement was negatively correlated with target prevalence. Their results indicated that at low prevalence, a higher, more conservative criterion was set, resulting in a greater number of misses and fewer false alarms. However, at high prevalence, a lower, more liberal criterion was set, resulting in fewer misses and more false alarms. These findings, therefore, suggest that shifting the observer's criterion towards being more liberal should make them

more likely to respond that there is a target, thus reducing the number of misses in low prevalence search tasks (Wolfe et al., 2007). Interventions that promote a less conservative criterion would be highly beneficial in real-world low prevalence settings, as they could reduce the number of undetected targets. In turn, this could prevent weapons and explosives from making it on to planes, and prevent cancers from going undetected.

Multiple-Decision Model

Wolfe and Van Wert (2010) propose a 'Multiple-Decision Model' to account for both the increased misses and speeded target-absent RT components of the LPE (see Figure 5). This is an adaption of a standard diffusion model. Such an adaptation is required, as a standard diffusion model requires changes to multiple parameters for it to align with the existing data, although making such adjustments would not fully account for the patterns observed in the data (Wolfe & Van Wert, 2010). According to this model, the observer begins by assessing some feature of the stimuli. If evidence exceeds the observer's decision criterion in the initial decision phase, a 'yes' response is given, indicating that a target is present, and the search is discontinued. A second decision phase is responsible for 'no' responses, which indicate that no target is present. This decision phase consists of a quitting threshold that can be reached via a diffusion process. Here a diffusion process refers to information accumulating over time, building towards the quitting threshold (Ratcliff & McKoon, 2008). If this quitting threshold is reached, a 'no' response is given and the search ends. If this threshold is not reached, the observer selects another feature of the stimuli and the process continues until a 'yes' or 'no' response is elicited.

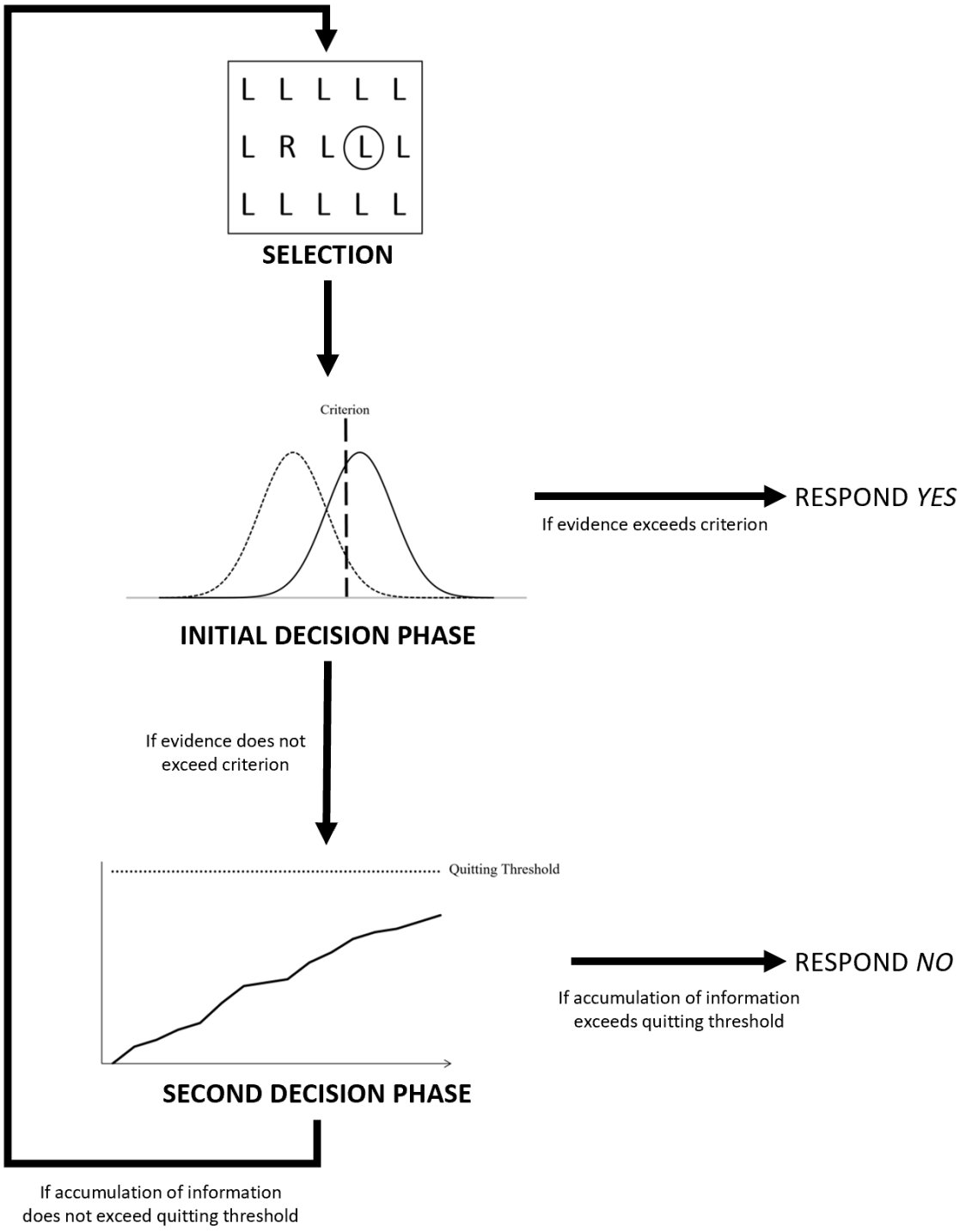


Figure 5. Visual depiction of the Multiple-Decision Model. Adapted from "Varying target prevalence reveals two dissociable decision criteria in visual search," by J. M. Wolfe, and M. J. Van Wert, 2010, *Current Biology, 20*, p.123.

According to this model both the decision criterion and the quitting threshold vary in response to target prevalence (Wolfe & Van Wert, 2010). At low prevalence, the criterion becomes more conservative, biasing the observer towards 'no' responses. The quitting threshold is also reduced at low prevalence, therefore, requiring less accumulation of information to terminate the search. This reduced quitting threshold results in speeded target-absent responses. Conversely, at high prevalence, the criterion becomes more liberal, biasing the observer towards 'yes' responses, and the quitting threshold is increased, consequently requiring a greater accumulation of information to terminate the search, which results in slower target-absent responses. Support for this model comes from a simulation which produced results that aligned with Wolfe and Van Wert's (2010) findings of speeded target-absent RTs and more conservative criterion placement at low prevalence, and slowed target-absent RTs and less conservative criterion placement at high prevalence.

Overcoming the Low Prevalence Effect

Increasing the Number of Targets

As mentioned earlier, the fact that a more liberal criterion placement results in fewer misses (Green & Swets, 1966; Wolfe & Van Wert, 2010), shifting the observer's criterion towards being more liberal should make them more likely to respond that there is a target, therefore, reducing the number of misses (Wolfe et al., 2007). One potential approach to achieve this is to increase target prevalence by increasing the number of target-present trials (Wolfe et al., 2007). This approach is already somewhat implemented by a quality control measure utilised by airport security known as 'Threat Image Projection' (Wolfe et al., 2013). This involves targets being projected into the x-ray luggage images at airport checkpoints (Wolfe et al., 2013). However, such an approach raises concerns due to the *Satisfaction of*

Search phenomenon, which suggests that a true target is more likely to go undetected when a false target is projected into that image (Fleck, Samei, & Mitroff, 2010). This is because if the observer finds the false target first, they are likely to be satisfied that they have found a target and finish the search without continuing to look for further true targets.

Increasing the Perceived Number of Targets with Feedback

An alternative approach to overcoming increased misses in low prevalence settings is to manipulate the observer's perceived prevalence (Schwark et al., 2012). This can be achieved by providing explicit feedback (Schwark et al., 2012). Previous work has suggested that feedback is required for observers to learn to adjust their criterion to better align with the rate of targets in the task (Estes & Maddox, 1995). Studies providing observers with full feedback have supported this. In their study imitating airport luggage screening, Wolfe et al., (2007) found that by implementing brief, higher prevalence retraining blocks that provided observers with feedback on every trial (see Figure 6), they were able to induce a less conservative criterion, which reduced misses during the low prevalence blocks. Providing brief retraining intervals throughout the low prevalence search task is thought to enable observers to recalibrate the prevalence rate at these higher prevalence points, which would then encourage a less conservative criterion that aligns with the higher target prevalence rate to be set and maintained across the low prevalence blocks (Wolfe et al., 2007). It is suggested that providing feedback in these higher prevalence retraining blocks is necessary, as without feedback, observers may not notice the higher prevalence rate (Wolfe et al., 2013). The findings from this retraining approach have also been replicated among newly trained TSOs (Wolfe et al., 2013).

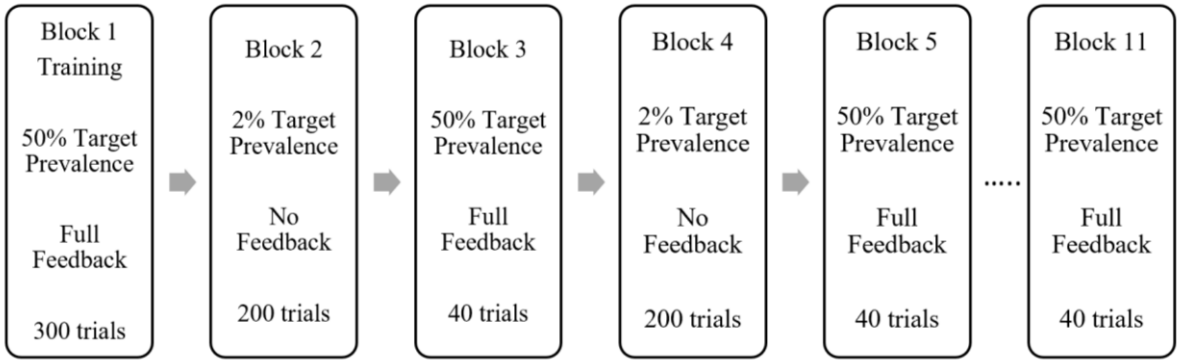


Figure 6. Visual representation of the method used in Wolfe et al.'s (2007) study.

Further evidence that feedback can be used to overcome the LPE comes from Schwark et al.'s (2012) study which used false feedback to shift observers' criterion. On 20% of trials in Schwark et al.'s (2012) study, participants were incorrectly informed that they had missed a target, leading them to believe that there were a greater number of targets than there actually were. The results indicate that false feedback shifted observer's criterion in a more liberal direction, resulting in those who received false feedback identifying significantly more low prevalence targets than those who received true feedback. These findings, therefore, also support the notion that increasing the observer's perceived number of targets can influence their criterion placement. However, it is worth noting that this approach of providing false feedback to observers to shift their criterion is not ideal ethically, as it involves lying to them about the accuracy of their performance, and would therefore, risk undermining the trust of those working in these low prevalence domains if used in real-world settings (Schwark et al., 2012).

The Current Study

The current study aims to explore the effectiveness of an alternative retraining procedure to that used by Wolfe et al. (2007). This retraining procedure is

based on selective feedback (i.e. observers receive correct feedback when they miss a target, but receive no feedback for hits, correct rejections, or false alarms), rather than full feedback. There is support in the recognition memory literature for the notion that feedback may only be required for specific response types, rather than for every response, to shift criterion. For instance, Han and Dobbins (2009) discovered that when true feedback was provided, but omitted for false alarms (i.e. indicating that a previously seen word was present, when it was not), individuals' decision criteria were more liberal. Conversely, omitting feedback for misses (i.e. failing to recognise a previously seen word when it is present) resulted in more conservative criteria. These shifts in criterion were also maintained once feedback was removed. This suggests that feedback for every response option is not necessary to induce and maintain a shift in criterion.

There is also reason to believe that selective feedback may be more effective at shifting criterion than full feedback. In their study investigating the impact of selective feedback on decision-making behaviour, Elwin, Juslin, Olsson, and Enkvit (2007) determined that participants' decision-making behaviour aligned with a constructivist coding approach. Constructivist coding occurs when an individual forms a mental representation of what they *believe* to be true, as opposed to what they *know* is true (Elwin et al., 2007). Therefore, according to the constructivist approach, when an individual receives external feedback about a decision they have made, they use this information to update their mental representation of the situation. However, when they do not receive external feedback, they assume that their decision is correct and also use this internal feedback to update their mental representation (Elwin et al., 2007). Therefore, in the context of a visual search task where observers only receive explicit feedback when they miss a target (i.e. selective

feedback), if the observer receives explicit feedback that they miss a target, this information is used to update their mental representation of the prevalence of targets. Conversely, if the observer receives no feedback when they correctly identify the target (i.e. hit) or correctly indicate that the target is absent (i.e. correct rejection), they will assume that they are correct and also use this information to update their mental representation of the prevalence of targets. However, if the observer incorrectly identifies a target when one is not present (i.e. false alarm) and they do not receive explicit feedback to correct them, they will also assume that they are correct in thinking that a target was present and use this information to update their mental representation of the prevalence of targets. Therefore, if participants use constructivist coding to mentally represent the prevalence of targets, selectively providing explicit feedback only when misses occur should increase the observer's perceived number of targets more so than when full feedback is provided, as this selective feedback will lead observers to believe that there are a greater number of targets than there actually are. This, in turn, should cause a greater liberal shift in criterion when feedback is only provided for misses than when full feedback is provided.

Aims and Hypotheses

Based on these findings, the current study aims to investigate the impact of selective feedback on criterion placement in visual search tasks. More specifically, the current study seeks to determine whether providing feedback for misses alone during brief higher prevalence retraining blocks leads to observers adopting a less conservative criterion that results in fewer misses, and more hits and false alarms. It also aims to determine whether selective feedback is more effective at inducing criterion shifts than full feedback.

Based on the notion that feedback in the retraining blocks should make the observer aware of the higher prevalence rate (Wolfe et al., 2013), it is hypothesised that full feedback will lead observers to maintain a less conservative criterion than when no feedback is provided. This, in turn, should result in a higher proportion of hits and false alarms in the full feedback condition than the control condition (Green & Swets, 1966). As discussed above, however, selective feedback in which observers only receive feedback when a target is missed is hypothesised to result in a greater liberal criterion shift than full feedback, and therefore, result in a greater proportion of hits and false alarms as well (Green & Swets, 1966).

Sensitivity is not expected to be impacted by feedback, as there is no reason to believe that either type of feedback should impact observers' ability to discriminate between the target and distractors. However, this outcome measure was included in this study in case, through some mechanism that has not been considered here, feedback does impact observers' ability to discriminate between the target and distractors. However, the results from statistical analyses on this measure should be interpreted tentatively, and potentially followed up with further research, as there are reasons to believe that the results generated from statistical tests on sensitivity may not be reliable in low prevalence settings (Wolfe et al., 2007). This is due to there being a much greater number of target-absent trials than target-present trials, making it not uncommon for extreme values to arise, such as hit rates of 1.00 or 0.00. Therefore, statistical tests on this data are likely to be unreliable as these extreme values greatly impact the d' measure of sensitivity (Wolfe, 2012).

The final hypotheses regard the predicted patterns of RTs. Based on Wolfe and Van Wert's (2010) Multiple-Decision Model, as well as Wolfe et al.'s (2005) finding that RTs on target-absent trials are linked with target prevalence rates, it is

hypothesised that if feedback leads to a more liberal shift in criterion, the observer's quitting threshold will increase, thus requiring a greater accumulation of information for them to respond that no target is present. This would, consequently, result in slower target-absent responses than when no feedback is provided. Conversely, target-present RTs are not expected to be impacted by feedback, as they have been shown to remain relatively stable regardless of target prevalence (Wolfe & Van Wert, 2010).

Method

Design

This study utilised a 3 (feedback condition) x 2 (target prevalence rate) mixed design. The feedback condition variable was between subjects, consisting of three levels: no feedback (control), full feedback (feedback on all retraining trials), and selective feedback (feedback for misses during the retraining blocks). The target prevalence rate variable was within subjects, consisting of two levels: 50% and 2.7% target prevalence. Separate analyses were conducted for criterion, hits, false alarms, target-present RTs, target-absent RTs, and sensitivity.

The desired sample size was 60 participants as this would align with Simmons, Nelson and Simonsohn's (2011) recommendation of a minimum of 20 participants per cell, which is suggested to ensure enough power to identify effects when effects are present. This is notably larger than samples used in previous studies (e.g. less than 20 in Wolfe et al., 2007).

Participants

The initial sample consisted of 57 participants. However, six sets of data were removed due to either being incomplete, confusion of the responding keys, containing blocks in which the large majority of responses were less than 500

milliseconds (indicating that the participant was not engaging with the task), or due to being a significant, or extreme, outlier on multiple measures in the first two blocks (suggesting that they did not understand the task).

The final sample, therefore, consisted of 51 participants (34 female) aged 18-50 years (*M* =26.57 years, *SD* =6.19 years) from the University of Tasmania and the greater Hobart community. Due to the nature of the visual stimuli used in the study, all participants were required to have normal or corrected-to-normal vision and were not colour blind. Random allocation of participants to feedback conditions resulted in 16 participants in the control condition, 17 in the selective feedback condition, and 18 in the full feedback condition. Participants received reimbursement for their time, either in the form of two hours of research credit for applicable students, or a $40 payment.

Materials

Test stimuli were presented on computer monitors using PsychoPy software. These stimuli consisted of coloured x-ray images of luggage which were obtained from the United States Department of Homeland Security (see Figure 7 for an example of the stimuli). These images consisted of an array of items that overlapped, with the degree of clutter varying across images. Target-absent arrays contained distractor items only, while target-present arrays contained distractor items and a single target item. The target item used in this study was a knife. The same target was used throughout the study to simplify the training required for participants to learn what a target item looked like in x-ray format, as participants had no prior experience with screening luggage. A knife was selected as the target over other targets in the airport security context, such as explosives, as less training is required to learn to detect this object (Wolfe et al., 2007). The knife was approximately 100

pixels long and 10 pixels wide. It was always positioned canonically, however, the location and orientation (0°, 45°, 90°, 135°, 180°, 225°, 270° or 315°) of the target varied across target-present arrays (see Figure 8 for the image of the knife used).

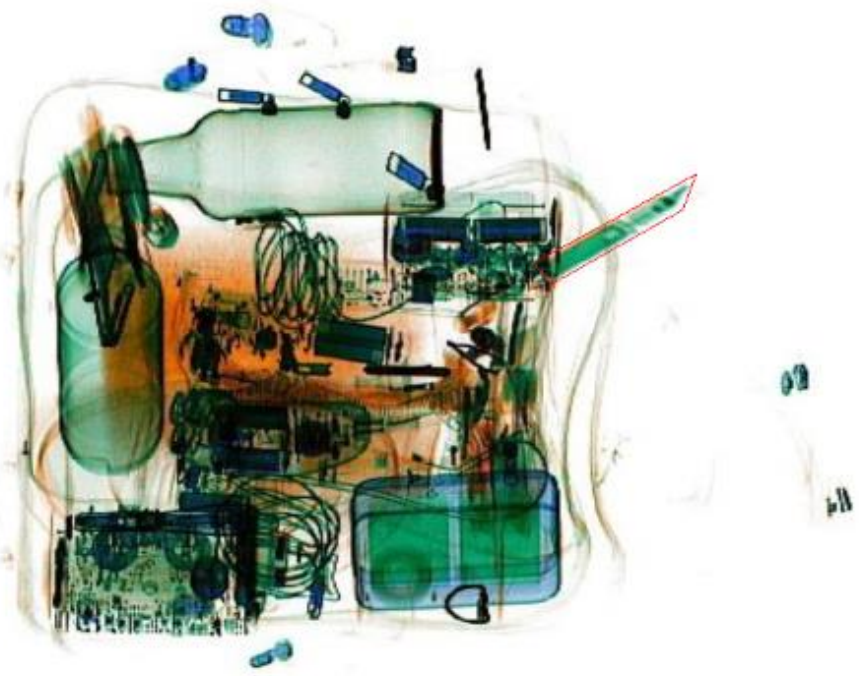


Figure 7. Example of test stimuli.

Figure 8. Example of target.

Procedure

The study commenced once ethics approval was granted (see Appendix A for the Ethics Approval letter). Participants were randomly allocated to one of the three feedback conditions. They completed the study individually at a computer in a laboratory. Participants were instructed to imagine that they were a luggage screener

at an airport security checkpoint where their task was to search x-ray images of luggage and decide whether a knife was present in each image. Participants were further informed that the task contained different phases. The first phase would be a *practice* phase, enabling them to become familiar with the task. The remainder of the study would consist of *training* and *on the job* phases. They were informed that on the job phases were similar to a security guard screening real bags. The information they received regarding the training phases varied depending on feedback condition. Participants in the control condition were informed that the training phase was similar to the practice phase, but that they should think of it as being the type of training an actual security guard would receive. Participants in the selective feedback condition were provided with the same information, with the addition that they *may* receive feedback about the accuracy of their response. Participants in the full feedback condition were also provided with the same information as the control condition, however, they were additionally informed that they *would* receive feedback about the accuracy of their response (see Appendix B for the instructions given to participants).

After reading the information sheet (see Appendix C) and providing consent (see Appendix D for the consent form) participants were presented with an example image of the target item in x-ray format (see Figure 8), and examples of target-present and target-absent arrays (see the images in Appendix B) to familiarise participants with the target and the upcoming task. When the study commenced, test stimuli were presented on the screen one at a time and remained there until the participant responded. The ‘D’ and ‘L’ keys were used for target-present and target-absent responses, and were counterbalanced between participants.

The method was similar to that used in Experiment 7 of Wolfe et al.'s (2007) study in that short retraining intervals were implemented amongst longer testing blocks. The current study consisted of eight blocks of trials which varied in terms of target prevalence, number of trials and availability of feedback (see Figure 9). The first block of the study was the practice phase which consisted of 40 no feedback trials at 50% prevalence. The remaining seven blocks alternated between on the job blocks (consisting of 185 no feedback trials at 2.7% prevalence) and training blocks (consisting of 40 trials at 50% prevalence and feedback that aligned with the participant's feedback condition). Different images were used in the practice phase and the three training phases, however, they all contained an equivalent degree of difficulty (the difficulty of images was estimated in a pilot of a previous study; see Bishop, 2014). The same target-absent images were used in each of the on the job phases but appeared in different orders. However, different target-present images were used in each block, and a target-absent version of these target-present images was not included in the practice or training phases.

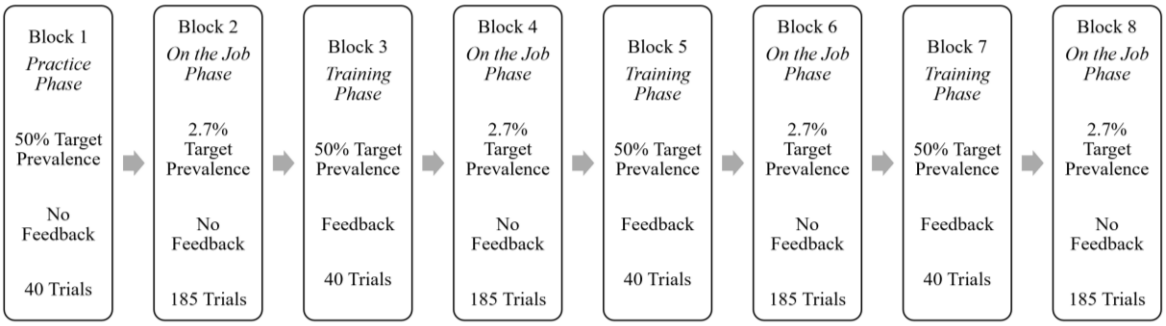


Figure 9. Visual representation of the method used in this study.

Feedback.

In the practice and on the job phases, participants received no feedback regardless of the condition they were assigned to. In the training phases, however, participants received feedback that aligned with their feedback condition. Feedback consisted of a message presented on the screen that reflected the accuracy of their response. This appeared immediately after a response was made and remained on the screen for two seconds before the next image appeared. When feedback was not provided for a trial in the training phases, a blank screen was presented for two seconds before the next image appeared to ensure the same duration of time between stimuli.

In the full feedback condition, feedback for correctly responding that a target was absent (i.e. a correct rejection) read, “Correct – there was no knife”. The feedback for correctly responding that a target was present (i.e. a hit) read, “Correct – you found the knife”. The feedback for incorrectly responding that a target was present (i.e. a false alarm) read, “Incorrect – you falsely identified the knife”. The feedback for incorrectly responding that a target was absent (i.e. a miss) read, “Incorrect – you missed the knife”.

In the selective feedback condition, feedback for incorrectly responding that a target was absent (i.e. a miss) read, “Incorrect – you missed the knife”. For all other response, participants received a blank screen with no feedback. Participants in control condition received a blank screen with no feedback for all responses

Realism of the task.

This design was utilised to imitate a real-world airport security setting. The long low prevalence blocks represent the low prevalence search tasks that take place at airport checkpoints, and the short higher prevalence blocks represent the retraining intervention that can be implemented into the low prevalence search task. The low

prevalence blocks that followed the retraining blocks (i.e. blocks 4, 6 and 8) enabled the identification of criterion shifts being maintained. It is worth noting that the prevalence rate in the low prevalence blocks does not necessarily align with that in the real-world setting, as the target prevalence rate in airport luggage is unknown (Wolfe et al., 2013). However, the targets in these blocks, much like in the real-world setting, were rare.

As TSOs are rotated to different tasks approximately every 20 minutes to reduce the impact of fatigue (Wolfe et al., 2007), breaks were also implemented at the end of the on the job phases. A minimum break of two minutes was enforced, however, participants were able to take longer if needed. In total, it took participants approximately 120 minutes to complete a total of 900 trials.

Results

Preliminary analyses included conducting a One-Way Analysis of Variance (ANOVA) for each of the outcome variables to determine whether the three feedback conditions differed on the first block of trials, prior to the feedback manipulation. The results indicated that feedback conditions did not significantly differ in terms of criterion placement, $F(2, 48) = 1.85$, $p = .168$, proportion of hits, $F(2, 48) = 2.10$, $p = .133$, proportion of false alarms, $F(2, 48) = 1.21$, $p = .308$, target-absent RTs, $F(2, 48) = 0.12$, $p = .887$, target-present RTs, $F(2, 48) = 0.00$, $p = .996$ or sensitivity, $F(2, 48) = 1.75$, $p = .185$ (see Appendix E for descriptive statistics). This, therefore, enables conclusions to be drawn about the effect of feedback on these outcome variables, as there were no significant differences between the three feedback groups prior to the manipulation taking place.

A 3 (feedback condition: control, selective feedback and full feedback) x 3 (on the job block: block 4, block 6 and block 8) mixed ANOVA was conducted for

each of the dependent variables. The first on the job block (block 2) was not included in these analyses as the manipulation did not take place until after this block. Significant interactions were followed up by splitting the data by feedback condition and running One-Way Repeated ANOVAs for the effect of block in each feedback condition. Significant ANOVAs were then followed up with pairwise comparisons that were Bonferroni adjusted to protect against Type I errors. Main effects were also followed up with Bonferroni adjusted pairwise comparisons. Results were interpreted by considering the effect size as well as statistical significance.

Where Mauchly's Test of Sphericity was significant, indicating a violation of the assumption of sphericity, a Greenhouse-Geisser correction was applied to the repeated factor, as well as to the interaction. Similarly, unless stated otherwise, Levene's Test of Equality of Error Variances was non-significant, indicating that the homogeneity of variance assumption was met.

Criterion

Criterion was calculated using the formula, c = 0.5[z(Hit rate)+z(False Alarm Rate)] (Macmillan & Creelman, 2005). There was a large, significant main effect of feedback condition on criterion, $F(2, 48) = 7.05$, $p = .002$, $\eta^2_p = .227$. Bonferroni adjusted pairwise comparisons (see Table 1) indicate that while the control condition demonstrated the most conservative criterion, followed by full feedback, then selective feedback, the only significant difference was between the control condition and selective feedback condition. However, as indicated by Cohen's *d* (Cohen, 1988), a moderate to large effect size was found for the difference between the selective and full feedback conditions, and a moderate effect was found for the difference between the control and full feedback conditions, therefore, suggesting meaningful differences between all feedback conditions.

Table 1

Bonferroni Adjusted Pairwise Comparisons of Criterion Across Feedback Conditions

Comparison	Mean Difference	*95% CI*		Cohen's *d*
		Lower	Upper	
Control vs. Selective Feedback	0.62*	0.21	1.04	1.29
Control vs. Full Feedback	0.24	-0.17	0.65	0.51
Selective Feedback vs. Full Feedback	-0.38	-0.78	0.02	0.79

Note. CI = Confidence Intervals; Positive mean values indicate that the first listed mean is greater than the second listed mean, while negative mean values indicate that the second listed mean is greater than the first listed mean.

*p <.05.

There was also a large, significant main effect of block on criterion, $F(2, 96) = 31.96$, $p < .001$, $\eta^2_p = .400$. However, this effect is better interpreted in the context of a large, significant interaction between feedback condition and block $F(4, 96) = 3.00$, $p = .022$, $\eta^2_p = .111$, which indicates that the variation in criterion across blocks is different across feedback conditions. Criterion varied across blocks in the control, $F(2, 30) = 24.07$, $p < .001$, $\eta^2_p = .616$, selective feedback, $F(2, 32) = 4.65$, $p = .017$, $\eta^2_p = .225$, and full feedback, $F(1.34, 22.73) = 14.37$, $p < .001$, $\eta^2_p = .458$ (following a Greenhouse-Geisser correction), conditions. As can be seen in Figure 10, in all three conditions criterion significantly increased, therefore, becoming more conservative, from block 4 to block 8. However, the control condition also significantly increased from block 4 to block 6, and the full feedback condition significantly increased from block 6 to block 8 (see table 2 for Bonferroni adjusted pairwise comparisons; note

that the assumption of homogeneity of variance was violated for block 8). Overall, this suggests that criterion becomes more conservative over time, but that this was reduced by feedback, especially selective feedback.

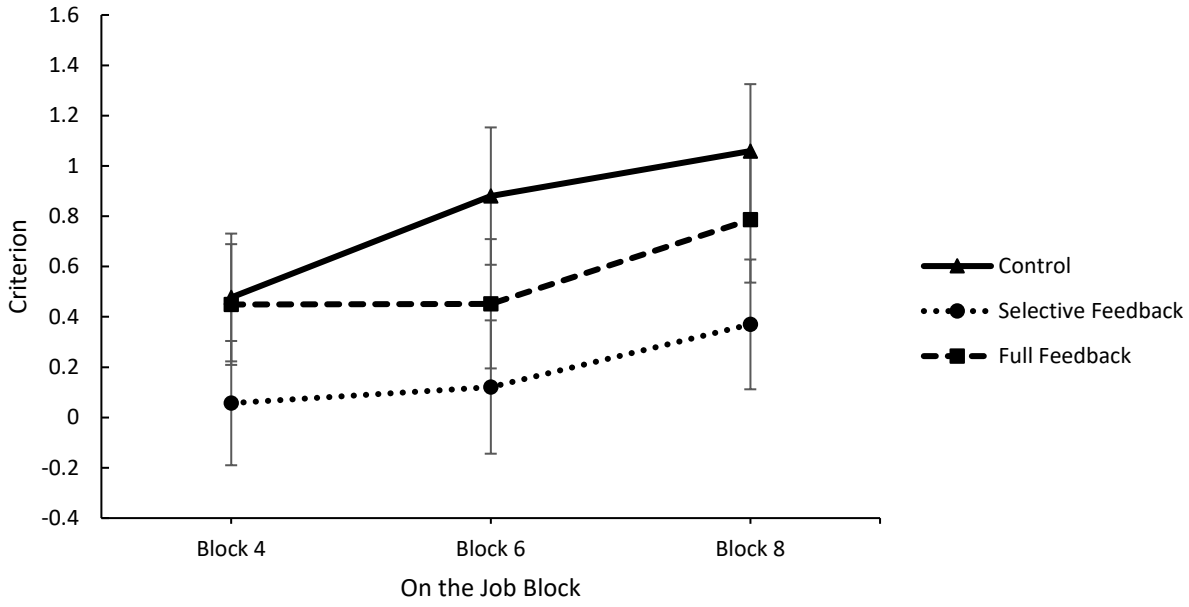


Figure 10. Estimated marginal means for criterion in each feedback condition across On the Job blocks. Error bars represent 95% confidence intervals.

Hits

There was a large, significant main effect of block on the proportion of hits, $F(2, 96) = 27.27$, $p < .001$, $\eta^2_p = .362$. However, this effect is better interpreted in the context of a large significant interaction between feedback condition and block, $F(4, 96) = 4.84$, $p = .001$, $\eta^2_p = .168$, which indicates that the variation in the proportion of hits across blocks is different across feedback conditions. The proportion of hits varied across block in the control, $F(2, 30) = 19.16$, $p < .001$, $\eta^2_p = .561$, selective feedback, $F(2, 32) = 6.00$, $p = .006$, $\eta^2_p = .273$, and full feedback, $F(2, 34) = 11.01$, $p < .001$, $\eta 2p = .393$, conditions. Bonferroni adjusted pairwise comparisons (see Table

Table 2

Bonferroni Adjusted Pairwise Comparisons of Criterion Across on the Job Block Per Feedback Condition

Comparison	Mean Difference	*95% CI*		Cohen's *d*
		Lower	Upper	
Control Condition				
Block 4 vs. Block 6	-0.40*	-0.62	-0.18	1.33
Block 4 vs. Block 8	-0.58**	-0.82	-0.34	1.66
Block 6 vs. Block 8	-0.18	-0.42	0.06	0.57
Selective Feedback Condition				
Block 4 vs. Block 6	-0.06	-0.34	0.22	0.15
Block 4 vs. Block 8	-0.31*	-0.59	-0.03	0.84
Block 6 vs. Block 8	-0.25	-0.56	0.06	0.54
Full Feedback Condition				
Block 4 vs. Block 6	0.00	-0.15	0.15	0.02
Block 4 vs. Block 8	-0.34*	-0.59	-0.09	0.94
Block 6 vs. Block 8	-0.33**	-0.49	-0.18	1.39

Note. CI = Confidence Intervals; Positive mean values indicate that the first listed mean is greater than the second listed mean, while negative mean values indicate that the second listed mean is greater than the first listed mean.

*p <.05; **p <.001.

3) indicate that the overall decrease from block 4 to block 8 was significant in the selective feedback and control conditions, as was the decrease from block 4 to 6 in the control condition, and the decrease from block 6 to block 8 in the selective and full feedback conditions.

Table 3

Bonferroni Adjusted Pairwise Comparisons of Hits Across on the Job Block Per Feedback Condition

Comparison	Mean Difference	95% CI		Cohen's *d*
		Lower	Upper	
Control Condition				
Block 4 vs. Block 6	1.31*	0.51	2.12	1.12
Block 4 vs. Block 8	1.81**	0.99	2.64	1.48
Block 6 vs. Block 8	0.50	-0.32	1.32	0.42
Selective Feedback Condition				
Block 4 vs. Block 6	0.06	-0.81	0.93	0.04
Block 4 vs. Block 8	0.94*	0.10	1.78	0.74
Block 6 vs. Block 8	0.88*	0.16	1.60	0.79
Full Feedback Condition				
Block 4 vs. Block 6	-0.33	-0.90	0.24	0.37
Block 4 vs. Block 8	0.72	-0.02	1.46	0.63
Block 6 vs. Block 8	1.06**	0.55	1.56	1.46

Note. CI = Confidence Intervals; Positive mean values indicate that the first listed mean is greater than the second listed mean, while negative mean values indicate that the second listed mean is greater than the first listed mean.

*p <.05; **p <.001.

The main effect of feedback on the proportion of hits was small and non-significant, $F(2, 48) = 1.18$, $p = .315$, $\eta^2_p = .047$, suggesting that the proportion of hits does not significantly vary across feedback conditions. However, inspection of effect

sizes for the difference between the means of the control and full feedback conditions, as well as for the difference between the means for the control and selective feedback conditions, at blocks 6 and 8 suggest that there are meaningful differences between the feedback conditions. Specifically, the effect size for the difference between the control and full feedback conditions was moderate at block 6 (d =0.63), but small at block 8 (d =0.26) with full feedback resulting in more hits in both blocks. For the difference between the control and selective feedback conditions, the effect size was large at block 6 (d =0.84), and moderate at block 8 (d =0.59), with selective feedback resulting in more hits in both blocks (see Figure 11). Overall, these results suggest that the proportion of hits decreases overtime, with feedback appearing to cause less of a reduction than no feedback.

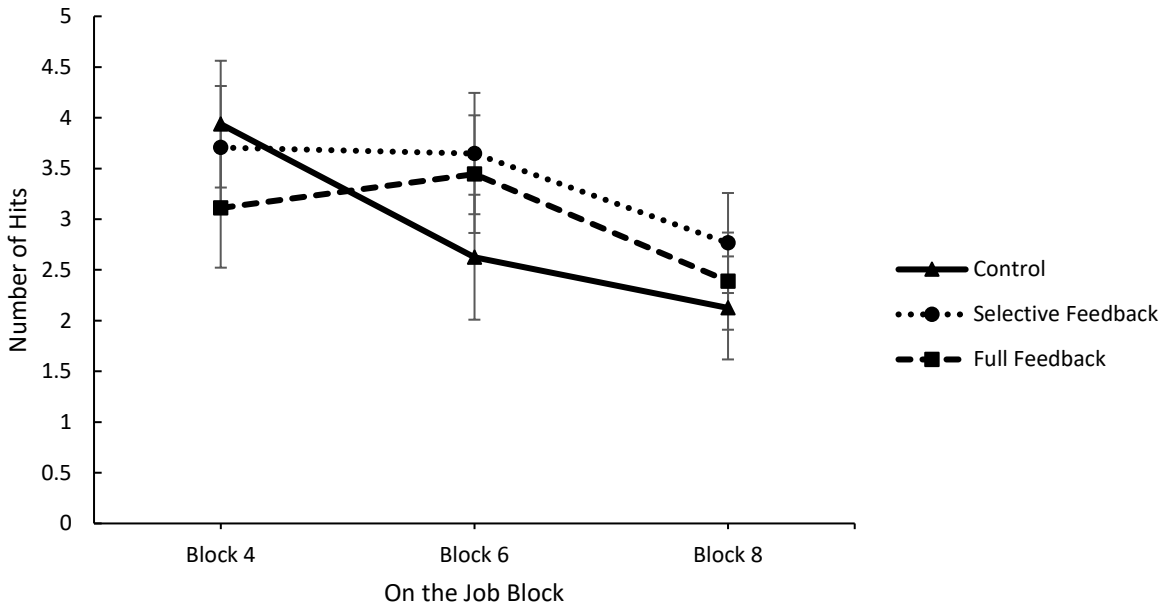


Figure 11. Estimated marginal means for the number of hits in each feedback condition across On the Job blocks. Error bars represent 95% confidence intervals.

False Alarms

A large, significant main effect of feedback condition on the proportion of false alarms was identified, $F(2, 48) = 7.91$, $p = .001$, $\eta^2_p = .248$. Bonferroni adjusted pairwise comparisons indicate that the proportion of false alarms was significantly greater in the selective feedback condition than in in the control and full feedback condition (as illustrated in figure 12), with both of these differences being a large effect (see table 4). The full feedback condition has the second highest proportion of false alarms, however, it was not significantly greater than the proportion of false alarms in the control condition.

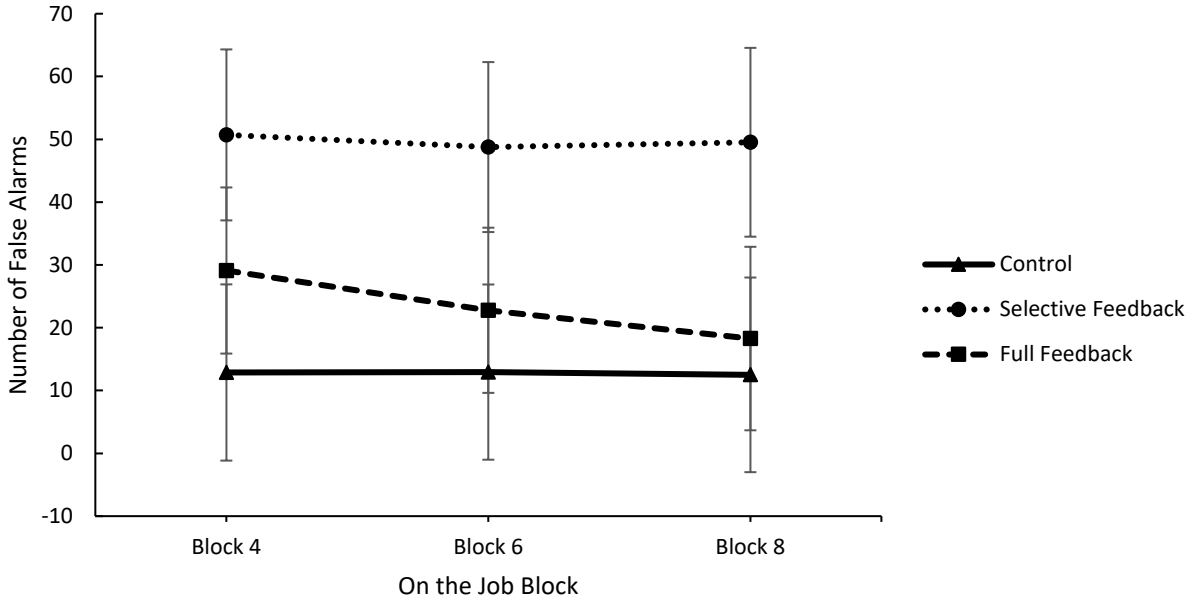


Figure 12. Estimated marginal means for the number of false alarms in each feedback condition across On the Job blocks. Error bars represent 95% confidence intervals.

Table 4

Bonferroni Adjusted Pairwise Comparisons of False Alarms Across Feedback Conditions

Comparison	Mean Difference	*95% CI*		Cohen's *d*
		Lower	Upper	
Control vs. Selective Feedback	-36.90*	-60.74	-13.05	1.34
Control vs. Full Feedback	-10.62	-34.14	12.90	0.38
Selective Feedback vs. Full Feedback	26.28*	3.13	49.43	0.95

Note. CI = Confidence Intervals; Positive mean values indicate that the first listed mean is greater than the second listed mean, while negative mean values indicate that the second listed mean is greater than the first listed mean.

*$p < .05$

The main effect of block on the proportion of false alarms following a Greenhouse-Geisser correction was small and non-significant, $F(1.62, 77.74) = 2.09$, $p = .139$, $\eta^2_p = .042$, therefore, indicating that the number of false alarms remains relatively stable across blocks. There was also a medium, but non-significant interaction between feedback condition and block on the proportion of false alarms following a Greenhouse-Geisser correction, $F(3.24, 77.74) = 1.41$, $p = .244$, $\eta^2_p = .056$, indicating that the effect of the feedback conditions on the proportion of false alarms did not differ significantly across all three blocks (note that the assumption of homogeneity of variance was violated for all three blocks). Overall, these findings indicate that selective feedback results in a greater proportion of false alarms than full feedback and no feedback, and that this effect is maintained across all three blocks.

The results discussed this far indicate that selective feedback successfully leads to a less conservative criterion than when full feedback of no feedback is provided. As a result, a much greater proportion of false alarms was observed for those receiving selective feedback, as was a slight increase in hits, compared to the control condition.

Target-Present Response Time

While it is understood that low prevalence influences criterion placement, which in turn impacts hit rates (Wolfe at al., 2007), the Multiple-Decision Model suggests that low prevalence may also influence observers' hit rates by reducing their quitting threshold, which leads them to terminate the search too quickly (Wolfe and Van Wert, 2010). Therefore, RTs were also investigated in this study.

The main effect of feedback on target-present RTs was trivial and non-significant, $F(2, 48) = 0.10$, $p = .908$, $\eta^2_p = .004$, indicating that target-present RTs do not differ between feedback conditions. There was, however, a large significant main effect of block on target-present RTs following a Green-house-Geisser correction, $F(1.28, 61.63) = 17.67$, $p < .001$, $\eta^2_p = .269$. Bonferroni adjusted pairwise comparisons (see Table 5) indicate that target-present RTs become faster over time, with target-present RTs significantly decreasing with each block. This pattern can be observed in Figure 13. The interaction between feedback condition and block for target-present RTs following a Greenhouse-Geisser correction was trivial and non-significant, $F(2.57, 61.63) = 0.40$, $p = .720$, $\eta^2_p = .017$, indicating that the pattern of decreasing target-present RTs over blocks is consistent across all three feedback conditions. Overall, these findings indicate that regardless of feedback type, target-present RTs become faster over time.

Table 5

Bonferroni Adjusted Pairwise Comparisons of Target-Present RT Across on the Job Blocks

Comparison	Mean Difference	95% CI		Cohen's *d*
		Lower	Upper	
Block 4 vs. Block 6	1.03*	0.22	1.85	0.48
Block 4 vs. Block 8	1.75**	0.85	2.65	0.76
Block 6 vs. Block 8	0.72**	0.34	1.10	0.69

Note. CI = Confidence Intervals; Positive mean values indicate that the first listed mean is greater than the second listed mean

*p <.05; **p <.001.

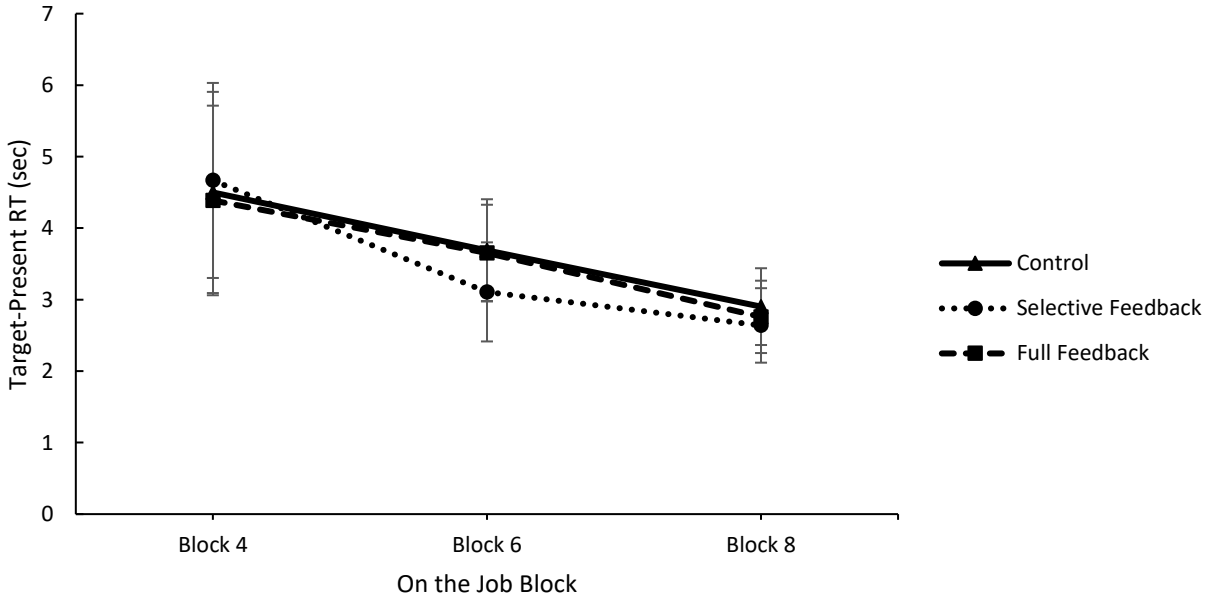


Figure 13. Estimated marginal means for target-present response times in each feedback condition across On the Job blocks. Error bars represent 95% confidence intervals.

Target-Absent Response Time

The main effect of feedback condition on target-absent RTs was trivial and non-significant, $F(2, 48) = 0.37$, $p = .690$, $\eta^2_p = .015$, indicating that target-absent RTs do not differ between feedback conditions. The main effect of block on target-absent RTs, however, was large and significant following a Greenhouse-Geisser correction, $F(1.23, 59.18) = 36.98$, $p < .001$, $\eta^2_p = .435$. Bonferroni adjusted pairwise comparisons (see table 6) indicate that target-absent RTs become faster over time, with target-absent RTs significantly decreasing with each block. This pattern can be seen in Figure 14. The interaction between feedback condition and block for target-absent RTs was trivial and non-significant following a Greenhouse-Geisser correction, $F(2.47, 59.18) = 0.29$, $p = .794$, $\eta^2_p = .012$, indicating that the pattern of decreasing target-absent RTs over blocks is consistent across all three feedback conditions. Overall, these results suggest that regardless of feedback type, target-absent RTs become faster over time. This suggests that feedback has no impact on observer's quitting thresholds. Therefore, taking all the findings discussed so far into account, it appears that selective feedback improves observers' hit rates by shifting their criteria, rather than by shifting their quitting threshold to make them search longer before quitting.

Sensitivity

As it was noted earlier, there are reasons to believe that results from statistical tests on sensitivity may be unreliable in low prevalence settings (Wolfe et al., 2007). Therefore, these results should be interpreted cautiously. For sensitivity, d' was calculated using the formula, d'= z(Hit rate)-z(False alarm rate) (Macmillan & Creelman, 2005). A large, but non-significant main effect of feedback condition on sensitivity was identified, $F(2, 48) = 2.84$, $p = .068$, $\eta^2_p = .106$, indicating that

Table 6

Bonferroni Adjusted Pairwise Comparisons of Target-Absent RT Across on the Job Blocks

Comparison	Mean Difference	95% CI		Cohen's *d*
		Lower	Upper	
Block 4 vs. Block 6	1.77**	0.81	2.72	0.71
Block 4 vs. Block 8	2.96**	1.89	4.02	1.09
Block 6 vs. Block 8	1.19**	0.78	1.60	1.13

Note. CI = Confidence Intervals; Positive mean values indicate that the first listed mean is greater than the second listed mean

***p* <.001.

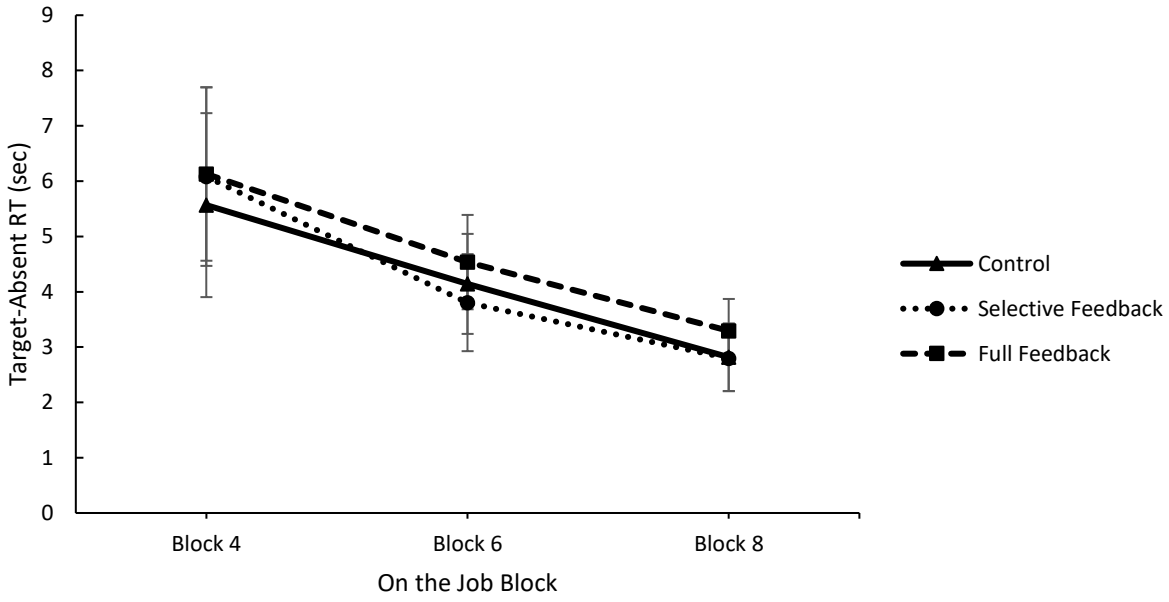


Figure 14. Estimated marginal means for target-absent response times in each feedback condition across On the Job blocks. Error bars represent 95% confidence intervals.

sensitivity does not differ significantly between feedback conditions. A large, significant main effect of block on sensitivity was found, $F(2, 96) = 9.16$, $p < .001$, $\eta^2_p = .160$. However, this effect is better interpreted in the context of a moderate, significant interaction between feedback condition and block, $F(4, 96) = 3.65$, $p = .008$, $\eta^2_p = .132$, which indicates that the variation in sensitivity across blocks is different across feedback conditions. Sensitivity varies across blocks in the control, $F(2, 30) = 10.65$, $p < .001$, $\eta^2_p = .415$, and full feedback, $F(2, 34) = 4.57$, $p = .018$, $\eta^2_p = .212$, conditions, but not the selective feedback condition, $F(2, 32) = 2.54$, $p = .095$, $\eta^2_p = .137$. Bonferroni adjusted pairwise comparisons (see Table 7) indicate that the overall decline in sensitivity from block 4 to block 8 was significant along with the decline from block 4 to block 6 in the control condition, as was the decline from block 6 to block 8 in the full feedback condition. Overall, as depicted in Figure 15, sensitivity appears to decline over time with feedback, especially selective feedback, reducing this decline.

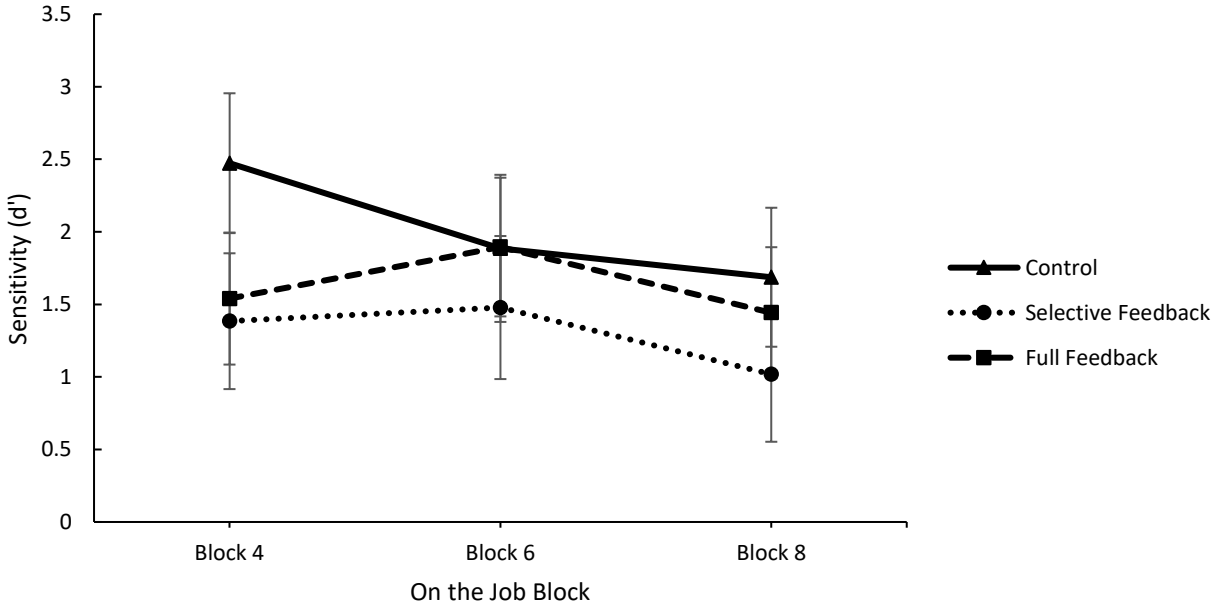


Figure 15. Estimated marginal means for sensitivity (d') in each feedback condition across On the Job blocks. Error bars represent 95% confidence intervals.

Table 7

Bonferroni Adjusted Pairwise Comparisons of Sensitivity Across on the Job Block Per Feedback Condition

Comparison	Mean Difference	*95% CI*		Cohen's *d*
		Lower	Upper	
Control Condition				
Block 4 vs. Block 6	0.59*	0.13	1.05	0.87
Block 4 vs. Block 8	0.79*	0.27	1.30	1.13
Block 6 vs. Block 8	0.20	-0.25	0.65	0.31
Selective Feedback Condition				
Block 4 vs. Block 6	-0.09	-0.60	0.41	0.12
Block 4 vs. Block 8	0.37	-0.30	1.03	0.37
Block 6 vs. Block 8	0.46	-0.09	1.01	0.55
Full Feedback Condition				
Block 4 vs. Block 6	-0.36	-0.78	0.06	0.55
Block 4 vs. Block 8	0.10	-0.39	0.59	0.13
Block 6 vs. Block 8	0.45*	0.12	0.79	0.90

Note. CI = Confidence Intervals; Positive mean values indicate that the first listed mean is greater than the second listed mean, while negative mean values indicate that the second listed mean is greater than the first listed mean.

*p <.05.

Discussion

The current study investigated whether selective feedback could cause a less conservative criterion, and therefore reduce misses, in low prevalence settings. Of particular interest was whether selective feedback was more effective at reducing

misses than full feedback, a method that has proven to be successful (Wolfe et al., 2007; Wolfe et al., 2013). Recognition memory studies have successfully induced less conservative criteria with the use of selective feedback (e.g. Han & Dobbins, 2009). This study, therefore, aimed to determine if such findings would extend to the domain of visual search.

Criterion Placement

The hypothesis that full feedback would lead to a less conservative criterion than no feedback is somewhat supported by the results of the current study. Although full feedback did not result in a significantly less conservative criterion than no feedback, there was a moderate effect size for the difference between these two conditions, suggesting that full feedback resulted in a meaningfully less conservative criterion than no feedback. This pattern aligns with Wolfe et al.'s (2007; 2013) findings that implementing brief higher prevalence retraining intervals containing feedback on every trial, results in a less conservative criterion during low prevalence, than when no feedback intervention is applied.

A larger, significant decrease in criterion for full feedback, compared to no feedback, may have not been observed in this study due to criterion in the first half of the study being reasonably neutral, especially compared to criteria that have been identified in low prevalence blocks of other studies (e.g. Wolfe et al., 2007). Therefore, introducing feedback that is designed to provide the observer with a neutral criterion would not result in a large shift, as their criterion was already set close to this, leaving little room for their criterion to shift. This early criterion placement is less likely to be an issue for selective feedback, as, depending on how many false alarms an observer makes, their criterion can be encouraged to become more liberal, not just neutral, thus leaving greater room for their criterion to shift.

The findings that selective feedback resulted in a significantly less conservative criterion than no feedback and that the difference between selective and full feedback had a moderate to strong effect size (despite being non-significant), supports the hypothesis that selective feedback would result in a less conservative criterion than both no and full feedback. These less conservative criteria that occur when feedback is provided are likely due to the feedback in the higher prevalence blocks making observers aware of the higher prevalence rate (Wolfe et al., 2013). Being aware of this would encourage observers to generate a criterion that is consistent with this higher prevalence rate (Wolfe et al., 2007). The criterion is then likely being maintained into the low prevalence blocks as these blocks do not provide feedback. Therefore, observers are not made aware of the lower prevalence rate, and are therefore, not recalibrating their criterion. Evidence for this being the mechanism through which less conservative criteria are maintained comes from Experiment 6 of Wolfe et al.'s (2007) study, which found that when using a similar intervention to the one used in this study, but with full feedback on both the low and high prevalence blocks, the criterion shifts were not maintained.

The finding that selective feedback leads to a meaningfully less conservative criterion than full feedback suggests that observers were processing feedback in a way that is consistent with constructivist coding (i.e. assuming that they are correct on hits, correct rejections, and false alarms as they did not receive feedback; Elwin et al., 2007), which led them to believe that target prevalence was higher than it actually was, and set a more liberal criterion to match this. This also aligns with Han and Dobbins' (2009) finding in recognition memory that omitting feedback that corrects false alarms leads to a more liberal criterion.

Criterion was also found to become increasingly conservative over time. Although this was not predicted, it is not an unusual finding as similar trends have been identified in recognition memory studies. For instance, Starns, Hicks & Marsh's (2006) study on word memory, which much like the current study involved a repetitious two-alternative forced choice task, also found an increase in criterion over time. Therefore, is seems reasonable that this pattern would emerge in the current study.

Overall, these findings suggest that implementing higher prevalence blocks with feedback into a low prevalence search task leads observers to maintain less conservative criteria. They also suggest that utilising selective feedback leads to an even lesser conservative criterion than when full feedback is utilised. Based on SDT, it would therefore be expected that these less conservative shifts in criterion should result in a greater proportion of hits and false alarms, with the proportions being greatest for the least conservative criterion (which here occurs for selective feedback; see Figure 4; Green & Swets, 1966).

Proportions of hits and false alarms.

As expected, the proportion of false alarms were strongly and significantly greater for selective feedback than for both no feedback and full feedback. The proportion of false alarms in the full feedback condition, however, was not meaningfully greater than the proportion in the control condition. However, this seems reasonable as criterion for full feedback was only moderately and non-significantly less conservative than for no feedback, therefore, a large increase in false alarms would not be expected.

It is worth noting that in some real-world situations, the increase in false alarms caused by selective feedback in this study would be concerning. For example,

in eyewitness identification, rejecting a line-up containing the culprit (i.e. a miss) enables the guilty individual to go free and potentially reoffend, but incorrectly identifying an innocent suspect (i.e. a false alarm) also has serious consequences as it leads to wrongful convictions (Clark & Godfrey, 2009; Rattner, 1988). However, in airport luggage screening, the consequence of a false alarm (e.g. holding up the line at an airport checkpoint) is not as severe as the consequence of a miss (e.g. life-threatening situations arising from weapons or explosives making it onto a plane). Therefore, while in some settings this increase in false alarms would be concerning, in the setting of airport luggage screening, where the consequences of a miss are far greater than the consequences of a false alarm, this increase in false alarms seems justified, provided that there is also a reduction in misses.

In terms of the proportion of hits, despite the unexpected finding that neither selective of full feedback resulted in a significant increase compared to no feedback, the moderate and large effect sizes for the differences between selective feedback and no feedback indicates that selective feedback resulted in a meaningfully greater proportion of hits than no feedback. However, full feedback only resulted in small and moderate differences from no feedback, which again seems reasonable based on the moderate, non-significant difference in criterion between full feedback and no feedback.

It seems likely that the reason a significant increase in false alarms was observed, but not a significant increase in hits, was due to the difference in opportunities to make false alarms and hits. The low prevalence blocks only contained five target-present trials. Therefore, as observers were already correctly identifying an average of 3.76 of the five targets prior to the manipulation, there was little room for improvement. Whereas, observers were making an average of 32.64

false alarms prior to the manipulation, but there was considerable room for improvement as the low prevalence blocks contained 180 target-absent trials. Therefore, it is likely that the small number of target-present trials, combined with the reasonably neutral criterion prior to the manipulation, as discussed above, has limited this study's ability to assess whether this feedback intervention leads to an increased proportion of hits.

Overall, these results suggest that the shift in criterion that results from implementing higher prevalence retraining blocks of selective feedback effectively increases false alarms compared to when full feedback or no feedback is used, but is not as effective at increasing hits. However, further research is required to confirm the effect of selective feedback on hits, as this finding may be impacted by a limitation of the current study's methodology.

Response Time

The hypothesis that if a less conservative criterion was observed, the observer's quitting threshold would increase, was not supported, as although differences in criterion placement were observed, feedback condition had no impact on target-absent RTs. This finding does not align with the quitting threshold component of the Multiple-Decision Model, as this predicts that at higher prevalence rates, the quitting threshold should increase, therefore requiring the accumulation of more information, which increases target-absent RTs (Wolfe & Van Wert, 2010). Therefore, as a higher prevalence rate is being perceived, as indicated by the observed shift in criterion, the quitting threshold should also increase. However, this result does somewhat align with Schwark et al.'s (2012) finding that despite false feedback resulting in a more neutral criterion at low prevalence than true feedback, target-absent RTs did not differ between the two conditions.

A potential explanation for this absence of increased target-absent RTs relates to a lack of motivation in the current sample to find the target. The more motivated an individual is to identify the target, the longer they will spend looking for it before terminating the search (Wolfe, 2012). This is well demonstrated in Wolfe's (2012) example that we are likely to be more motivated to find a missing $100 note than a $5 note, and are therefore likely to spend more time searching for a $100 note than a $5 note. Therefore, perhaps an increase in target-absent RTs did not occur due to participants not being motivation to locate the target. This seems plausible as there were no real-life gains to be achieved from identifying the knife, nor were there consequences for missing it, therefore, suggesting that there was minimal motivation for participants to identify the target. This brings to question whether these findings regarding target-absent RTs would be replicated by TSOs as these individuals would likely possess greater motivation to identify targets due to the severe consequences of missing a target. However, further research would be required to determine whether a lack of motivation is responsible for the absence of an increase in target-absent RTs in response to feedback.

As expected, target-present RTs did not differ across feedback conditions. As previous similar studies have not reported target-present RTs (e.g. Schwark et al., 2012; Wolfe et al., 2007; Wolfe et al., 2013), this data cannot be compared. However, this does align with Wolfe and Van Wert's (2010) finding that target-present RTs remained relatively stable regardless of variations in criterion. This indicates that this is not an unusual finding.

Sensitivity

The finding that sensitivity declines over time, with the degree of decline varying across feedback conditions was unexpected, as sensitivity has typically

remained relatively stable in previous studies (e.g. Wolfe et al., 2007). This suggests that observers' ability to discriminate between distractors and the target becomes worse over time, with declines in ability occurring in the control and no feedback condition. While this appears to be an unusual result, as it has been previously discussed, statistical tests on sensitivity can be unreliable in low prevalence settings due to the greatly unequal proportions of target-present and target-absent trials (Wolfe et al., 2007; Wolfe, 2012). Therefore, little emphasis should be placed on this finding, as further research is needed to clarify it.

Summary and Implications

These results provide valuable information regarding the Multiple-Decision Model, as well as the effectiveness of the selective feedback retraining intervention on visual search performance. Based on the results of the current study, it appears that implementing brief higher prevalence retraining blocks with selective feedback into low prevalence search tasks causes a less conservative criterion, resulting in a greater proportion of false alarms and potentially an increase in hits. Selective feedback also appears to be more effective than full feedback at achieving this. However, both selective and full feedback appear to have no impact on the quitting threshold as indicated by the lack of change in target-absent RTs (Wolfe and Van Wert, 2010). Therefore, both forms of feedback are ineffective at encouraging observers to search longer before terminating the search. This suggests that the observed increase in false alarms and hits are due to feedback causing a liberal decision criterion shift during the initial decision phase, making observers less biased towards responding that no target is present (Macmillan & Creelman, 2005), rather than shifting the quitting threshold. These findings suggest that introducing this selective feedback retraining intervention during low prevalence search tasks, such as

airport luggage screening, could reduce observers' biases towards responding that no target is present, and therefore, reduce the number of targets that go undetected. This is a valuable finding, as it seems reasonable that such an intervention could realistically be implemented into the task of routine airport luggage screening. Being able to shift airport security workers' response biases so that they are more willing to identify low prevalence targets, such as weapons and explosives, would ideally reduce the number of these targets going undetected. This would be highly valuable due to the life-threatening consequences associated with missing such targets.

Suggestions for Future Research and Addressing Limitations

Due to the fact that the outcomes of the current study would be valuable in real-word low prevalence settings, such as in airport luggage screening, the clearest next step would be to investigate whether these findings can be replicated with TSOs, much like how Wolfe et al. (2013) replicated the findings of Wolfe et al.'s (2007) full feedback retraining procedure. Attempting to replicate these findings with this real-world sample would also address the potential limitation of a lack of motivation in the current study, as TSOs, or newly trained TSOs as used in Wolfe et al.'s (2013) study, are likely to be motivated to identify the target due to the consequences of missing it . However, these future studies should also aim to include a greater number of target-present trials in the low prevalence blocks than what was used in the current study in order to investigate whether selective feedback can more effectively increase hits, as the small number of target-present trials in low prevalence blocks in the current study limited the investigation of this.

It would also be worth investigating if these effects of the selective feedback intervention can be replicated with a more realistic visual search task in which observers search for multiple low prevalence targets, not just a single knife. The use

of a single knife may have enabled observers to search for a specific feature. However, previous studies suggest that visual search behaviour differs when an observer is searching for a single feature compared to when they are searching for an overall target (Rich et al., 2008). Therefore, further research would be required to determine if these effects of selective feedback can be replicated when the opportunity to rely on single feature detection is removed.

A further aspect that would be valuable to investigate in future studies is whether the suspected underlying mechanism for this intervention working is in fact responsible for these results. Therefore, future studies could also collect information regarding participants' perceived prevalence rates, as it is currently understood that it is feedback's influence on this perception that causes shifts in criterion (Schwark et al., 2012), however, previous studies in this domain have not directly assessed this.

Conclusions

In conclusion, the current study indicates that implementing higher prevalence retraining blocks in which feedback is only provided for misses, into a low prevalence visual search task, results in a less conservative criterion than when full feedback or no feedback is provided. Consequently, selective feedback results in a greater proportion of false alarms. Although no significant impact of feedback on the proportion of hits was identified, selective feedback did result in a meaningfully higher proportion of hits than no feedback. Therefore, further research is required to determine whether this intervention can be used to increase hits, as a methodological limitation of the current study may have restricted this from occurring. It was also found that this retraining intervention had no impact on target-present or target-absent RTs, indicating that this selective feedback retraining intervention increases false alarms and hits by causing a liberal shift of the decision criterion in the initial

decision phase of the Multiple-Decision Model. Further investigation of this intervention with a more realistic real-world sample (e.g. TSOs) and a greater number of target-present trials is required to establish whether this intervention could be used to reduce misses of low prevalence targets during real-world airport luggage screening.

References

Biggs, A. T., Adamo, S. H., & Mitroff, S. R. (2014). Rare, but obviously there: Effects of target frequency and salience on visual search accuracy. *Acta Psychologica, 152,* 158-165. Doi:10.1016/j.actpsy.2014.08.005

Bishop, C. (2014). *Effect of feedback in low prevalence visual search* (Unpublished honours thesis). University of Tasmania, Tasmania.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafo, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*(5), 365-376. Doi:10.1038/nrn3475

Clark, S. E., & Godfrey, R. D. (2009). Eyewitness identification evidence and innocence risk. *Psychometric Bulletin & Review, 16*(1), 22-42. Doi:10.3758/PBR.16.1.22

Cohen J. (1988). Statistical power analysis for the behavioral sciences. New York: Routledge Academic.

Elwin, E., Juslin, P., Olsson, H., & Enkvist, T. (2007). Constructivist coding. *Psychological Science, 18*(2), 105-110. Doi:10.1111/j.1467-9280.2007.01856.x

Estes, W. K., & Maddox, W. T. (1995). Interactions of stimulus attributes, base rates, and feedback recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(5), 1075-1095. Doi:10.1037/0278-7393.21.5.1075

Evans, K. K., Birdwell, R. L., & Wolfe, J. M. (2013). If you don't find it often, you often don't find it: Why some cancers are missed in breast cancer screening. *PloS One, 8*(5), 1-6. Doi:10.1371/journal.pone.0064366

Evans, K. K., Tambouret, R. H., Evered, A., Wilbur, D. C., & Wolfe, J. M. (2011). Prevalence of abnormalities influences cytologists' error rates in screening for cervical cancer. *Archives of Pathology & Laboratory Medicine, 135*(12), 1557-1560. Doi:10.5858/arpa.2010-0739-OA

Fleck, M. S., & Mitroff, S. R. (2007). Rare targets are rarely missed in correctable search. *Psychological Science, 18*(11), 943-947. Doi:10.1111/j.1467-9280.2007.02006.x

Fleck, M. S., Samei, E., & Mitroff, S. R. (2010). Generalized "satisfaction of search": Adverse influences on dual-target search accuracy. *Journal of Experimental Psychology: Applied, 16*(1), 60-71. Doi:10.1037/a0018629

Förster, J., Higgins, E. T., & Bianco, A. T. (2003). Speed/accuracy decisions in task performance: Built-in trade-off or separate strategic concerns? *Organizational Behavior and Human Decision Processes, 90*(1), 148-164. Doi:10.1016/S0749-5978(02)00509-5

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Han, S., & Dobbins, I. G. (2009). Regulating recognition decisions through incremental reinforcement learning. Psychonomic Bulletin & Review, 16(3), 469-474. Doi:10.3758/PBR.16.3.469

Harris, D. H. (2002). How to really improve airport security. *Ergonomics in Design, 10*(1), 17-22. Doi:10.1177/106480460201000104

Lau, J. S. H., & Huang, L. (2010). The prevalence effect is determined by past experience, not future prospects. *Vision Research,* 50(15), 1469-1474. Doi:10.1016/j.visres.2010.04.020

Lehman, C. D., Arao, R. F., Sprague, B. L., Lee, J. M., Buist, D. S. M., Kerlikowske, K., … Miglioretti, D. L. (2017). National performance benchmarks for modern screening digital mammography: Update from the Breast Cancer Surveillance Consortium. *Radiology, 283*(1), 49-58. Doi:0.1148/radiol.2016161174

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide*. New Jersey: Lawrence Elbaum Associates

Mitroff, S. R., & Biggs, A. T. (2014). The ultra-rare-item effect: Visual search for exceedingly rare items is highly susceptible to error. *Psychological Science, 25*(1), 284-289. Doi:10.1177/0956797613504221

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation, 30*(4), 873-922. Doi:10.1162/neco.2008.12-06-420

Rattner, A. (1988). Convicted but innocent: Wrongful conviction and the criminal justice system. *Law and Human Behavior, 12*(3), 283-293. Doi:10.1007/BF01044385

Rich, A. N., Kunar, M. A., Van Wert, M. J., Hidalgo-Sotelo, B., Horowitz, T. S., & Wolfe, J. M. (2008). Why do we miss rare targets? Exploring the boundaries of the low prevalence effect. *Journal of Vision, 8*(15), 1-17. Doi:10.1167/8.15.15

Schwark, J., Sandry, J., Macdonald, J., & Dolgov, I. (2012). False feedback increases detection of low-prevalence targets in visual search. *Attention, Perception, & Psychophysics, 74*(8), 1583-1589. Doi:10.3758/s13414-012-0354-4

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting

anything as significant. Psychological Science, 22(11), 1359-1366. Doi:10.1177/0956797611417632

Starns, J. J., Hicks, J. L., & Marsh, R. L. (2006). Repetition effects in associative false recognition: Theme-based criterion shifts are the exception, not the rule. *Memory, 14*(6), 742-761. Doi:10.1080/09658210600648514

Van Wert, M. J., Horowitz, T. S., & Wolfe, J. M. (2009). Even in correctable search, some types of rare targets are frequently missed. *Attention, Perception, & Psychophysics, 71*(3), 541-553. Doi:10.3758/APP.71.3.541

Wolfe, J. M. (2012). When do I quit? The search termination problem in visual search. *Nebraska Symposium on Motivation, 59*, 183-208. doi:10.1007/978-1-4614-4794-8_8

Wolfe, J. M., Brunelli, D. N., Rubinstein, J., & Horowitz, T. S. (2013). Prevalence effects in newly trained airport checkpoint screeners: Trained observers miss rare targets, too. *Journal of Vision, 13*(3), 1-9. Doi:10.1167/13.3.33

Wolfe, J. M., Horowitz, T. S., Kenner, N. M. (2005). Rare items often missed in visual searches: Errors in spotting key targets soar alarmingly if they appear only infrequently during screening. *Nature, 435*(7041), 439-440. Doi:10.1038/435439a

Wolfe, J. M., Horowitz, T. S., Van Wert, M. J., Kenner, N. M., Place, S. S., & Kibbi, N. (2007). Low target prevalence is a stubborn source of errors in visual search tasks. *Journal of Experimental Psychology: General, 136*(4), 623-638. Doi:10.1037/0096-3445.136.4.623

Wolfe, J. M., & Van Wert, M. J. (2010). Varying target prevalence reveals two dissociable decision criteria in visual search. *Current Biology, 20*(2), 121-124. Doi:10.1016/j.cub.2009.11.066

Appendices

Appendix A

Ethics Approval Letter

Dear Dr Palmer

Ethics Ref: H0012660

Title: Confidence in memory

This email is to confirm that the following amendment was approved by the Chair of the Tasmania Social Sciences Human Research Ethics Committee on 7/8/2019:

We request an extension to this project until 30 September 2019 We request addition of the following student researchers to the project: Ashten de Haan (ID 388881) Tegan Marston-Pattinson (227918). Could you please advise the following regarding the request to add the above students:

- Full name and title
- Date of Birth
- Student ID
- College / School
- Current level studying
- Email address

All committees operating under the Human Research Ethics Committee (Tasmania) Network are registered and required to comply with the National Statement on Ethical Conduct in Human Research (NHMRC 2007, updated May 2015).

Please be reminded that all ethical approvals granted are subject to conditions as required by the National Statement. A copy of the conditions of approval is available at http://www.utas.edu.au/research-admin/research-integrity-and-ethics-unit-rieu/human-ethics/human-research-ethics-review-process/managing-your-ethics-approved-projects

This email constitutes official approval. If your circumstances require a formal letter of amendment approval, please let us know.

If you have any questions, please contact SS.Ethics@utas.edu.au or 03 6226 2975.

Kind regards

Jude

Jude Vienna-Hallam

Executive Officer, Social Science HREC

Research Integrity and Ethics Unit I Research Division

University of Tasmania

Building 1, 1st Floor, 301 Sandy Bay Road

Hobart TAS 7001

Telephone: 03 6226 2608

www.utas.edu.au/research-admin/research-integrity-and-ethics-unit-rieu

Appendix B

Participant Instructions

In this experiment, we would like you to imagine you are a luggage screener at an airport. You will be shown x-ray images of luggage and your task is to decide whether or not each contains a knife.

Here is an example of the knife you are looking for. In some images you might only see part of the knife and the colour could appear darker or lighter.

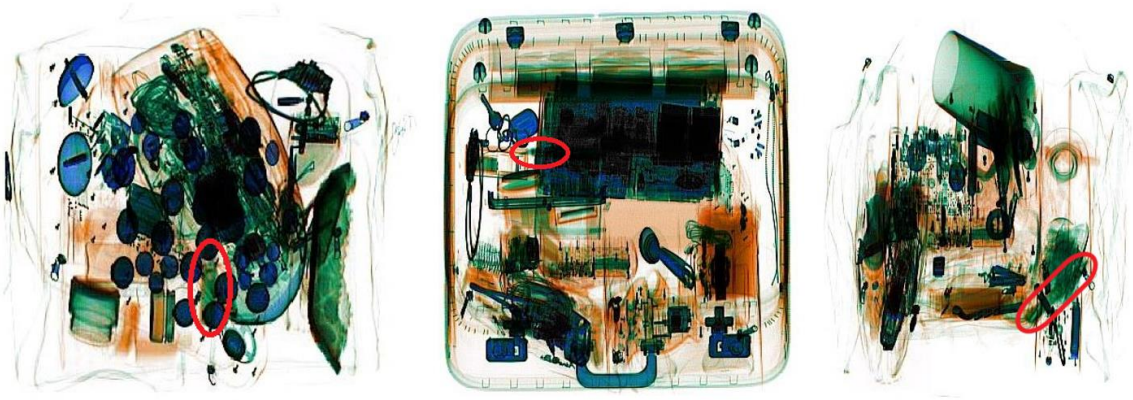
On the next screen you will see two example bags. One DOES NOT contain a knife, the other DOES contain a knife.

Bag DOES NOT contain knife

Bag DOES contain knife

Here are some more examples of the images you can expect to see.

Note that the knife may be partly obscured as shown below.



The experiment will be broken into different parts. The first part will be a "practice phase". You can think of this as a security screening training phase where you become familiar with the task. You will see some practice images and will decide whether each contains a knife.

If you think the bag DOES NOT contain a knife, press the ['D' or 'L' (depending on the participant's counterbalanced condition)] key.

If you think the bag DOES contain a knife, press the ['D' or 'L' (depending on the participant's counterbalanced condition)] key.

Please try to respond as quickly and accurately as possible.

{Following completion of the practice phase}.

The remainder of the experiment will consist of "training" and "on the job" phases.

On the job phase: Similar to the practice but think of it as being an actual security guard screening real bags.

Training phase {control condition}: Similar to the practice but think of this as the type of training an actual security guard would take part in.

Training phase {selective feedback condition}: Similar to the practice but you may receive feedback (e.g. whether your response was correct or not) on some trials. Think of this as the type of training an actual security guard would take part in.

Training phase {full feedback condition}: Similar to the practice but you will receive feedback (e.g. whether your response was correct or not} on all trials. Think of this as the type of training an actual security guard would take part in.

Appendix C

Information Sheet

Decision-Making in Luggage Screening

1. Invitation
You are invited to participate in a psychology experiment examining decision-making when screening security images of luggage. The study is being conducted by Honours student Ashten de Haan under the supervision of Dr Matthew Palmer, within the School of Psychology at the University of Tasmania.

2. What is the purpose of this study?
The experiment aims to investigate how people make decisions about whether an item of luggage contains a threat item.

3. Why have I been invited to participate?
You have been identified on the basis of being 18 years of age or older and a current student at the University of Tasmania or as a member of the wider community. Participation in this research is completely voluntary meaning you do not have to participate if you do not wish to and there will be no consequences. You are also free to leave the study at any time.

4. What will I be asked to do?
You will be asked to view several security-like images of luggage and determine whether or not each contains a threat item (i.e., a knife). Images will be shown individually on a computer screen. No prior knowledge or experience in luggage screening is required. The study will take approximately 2 hours to complete.

5. Are there any possible benefits from participation in this study?
There may be no direct benefits to yourself, however the information gained will provide key insight into psychological theories regarding the decisions made when screening luggage images for threats.

For their time, participants will receive either 2 research credits or a gift voucher for $40.00.

6. Are there any possible risks from participation in this study?
There are no foreseeable risks associated with participating in this research.

7. What if I change my mind during or after the study?
You are free to leave the study at any time without giving an explanation. However, once you have started the study it is not possible to withdraw as your information will be stored anonymously and therefore we cannot identify your particular responses.

8. What will happen to the information when this study is over?
The data from this study will be stored in a secure online service. When the data is ready to be analysed, it will be stored securely on the University of Tasmania's password protected server in a de-identified form (so your anonymity is preserved). The data will be kept for at least five years from the date of first publication. Whilst all data will be stored in a de-identified format, it will be at the researchers' discretion who to share this de-identified data with (e.g. other researchers upon request).

9. How will the results of the study be published?
The research findings will be reported in an academic journal. If you would like to access the final results please contact the researcher. No individual participants will be identified in the publication of this study.

10. What if I have questions about this study?
Should you have questions relating to any aspect of this research please feel free to contact Ashten de Haan (akde@utas.edu.au).

This study has been approved by the Tasmanian Social Sciences Human Research Ethics Committee. If you have concerns or complaints about the conduct of this study, please contact the Executive Officer of the HREC (Tasmania) Network on +61 3 6226 6254 or email human.ethics@utas.edu.au. The Executive Officer is the person nominated to receive complaints from research participants. Please quote ethics reference number **H0012660**.

If you wish to continue please press the space bar.

Appendix D

Consent Form

Decision-Making in Luggage Screening

UTAS

I agree to take part in the research study named above and understand the information previously provided for this experiment.

I understand that the study involves viewing security images of luggage and determining whether a threat item is present.

I understand that all research data will be securely stored on the University of Tasmania premises for at least five years from the publication of the study results.

Any questions that I have asked have been answered to my satisfaction.

I understand that the researchers will maintain confidentiality and that any information I supply to the researcher will be used only for the purposes of the research.

I understand that the results of the study will be published so that I cannot be identified as a participant.

I understand that my participation is voluntary and that I may withdraw at any time without any effect, however, will not be able to withdraw my data after completing the experiment as my data will be anonymous.

By pressing the space bar, you consent to the above conditions and to be part of this experiment.

Appendix E

Tables of Descriptive Statistics for the One-Way ANOVAs run on the Practice Phase Data

Table 8

Descriptive Statistics for Criterion for the Practice Phase

			95% CI	
Condition	Mean	SD	Lower	Upper
Control	0.27	0.39	0.06	0.47
Selective Feedback	0.01	0.37	-0.19	0.20
Full Feedback	0.17	0.41	-0.04	0.37

CI = Confidence Interval

Table 9

Descriptive Statistics for the Proportion of Hits for the Practice Phase

			95% CI	
Condition	Mean	SD	Lower	Upper
Control	14.19	2.32	12.95	15.42
Selective Feedback	15.06	2.86	13.59	16.53
Full Feedback	13.00	3.56	11.23	14.77

CI = Confidence Interval

Table 10

Descriptive Statistics for the Proportion of False Alarms for the Practice Phase

			95% CI	
Condition	Mean	SD	Lower	Upper
Control	3.44	3.10	1.79	5.09
Selective Feedback	5.06	3.15	3.44	6.68
Full Feedback	5.17	4.29	3.03	7.30

CI = Confidence Interval

Table 11

Descriptive Statistics for Target-Present RT for the Practice Phase

			95% CI	
Condition	Mean	SD	Lower	Upper
Control	5.48	2.68	4.05	6.91
Selective Feedback	5.42	2.38	4.19	6.64
Full Feedback	5.49	2.73	4.13	6.84

CI = Confidence Interval

Table 12

Descriptive Statistics for Target-Absent RT for the Practice Phase

			95% CI	
Condition	Mean	SD	Lower	Upper
Control	9.48	2.97	7.90	11.07
Selective Feedback	10.24	5.20	7.57	12.91
Full Feedback	9.74	4.93	7.29	12.19

CI = Confidence Interval

Table 13

Descriptive Statistics for Sensitivity for the Practice Phase

			95% CI	
Condition	Mean	SD	Lower	Upper
Control	1.70	0.75	1.30	2.10
Selective Feedback	1.50	0.68	1.15	1.85
Full Feedback	1.20	0.92	0.74	1.66

CI = Confidence Interval